

PAILA, HARI SRINIVAS KALYAN, M.S. Molecular Modeling Studies of Curcumin Analogs as Anti-Angiogenic Agents. (2008)
Directed by Dr. J. Phillip Bowen. 97 pp.

Curcumin is a carotenoid natural product isolated from the rhizome of the plant *Curcuma longa*. Among its many biological effects, curcumin has anti-inflammatory, anti-infective, and anti-cancer activity. The anti-cancer activity of curcumin has been studied extensively.

Angiogenesis plays a pivotal role in the metastasis of cancer: curcumin showed excellent anti-angiogenesis activity on metastatic tumors. Several curcumin analogues have been synthesized and studied, and their biological activity was reported in the literature. One class of potent analogues are aromatic enones. In Dr Bowen's laboratory sixty three compounds were synthesized and in the laboratory of Dr Jack Arbizer (Emory University, Atlanta, GA) they were tested for their anti-angiogenic activity with an SVR endothelial cell growth assay developed by Dr Arbizer. The precise mechanism or the specific biological target on which these analogs exert their inhibition potential as anti-angiogenic agents is unknown. Therefore, structure-based molecular modeling is not a possibility. However, ligand based molecular modeling methods are available for studying and predicting which compounds among the sixty three can be further optimized for selectivity and desired property.

Computational studies were carried out to identify which structural features within the series of analogues are significantly important for activity. Initially, pharmacophore modeling was carried out in Molecular Operating Environment (MOE) software to identify the Interaction Pharmacophore Elements (IPE) and their relative geometry in three-dimensional space. Two different three dimensional quantitative structural Activity Relationship (3D-QSAR) studies, Comparative Molecular Field Analysis (CoMFA), and Comparative Molecular Similarity Indices Analysis (CoMSIA) were carried out with this dataset. SYBYL (versions 7.2 and 7.3) were used for the development of the models. Forty six compounds were used as the calibration or the training set. The model yielded a cross validated q^2 of 0.289 for CoMFA and 0.146 for CoMSIA analyses. Eleven compounds were used as the test set (or the prediction) set to externally validate the QSAR models and their robustness. The predictions of the model are acceptable with a few outliers.

MOLECULAR MODELING STUDIES OF CURCUMIN ANALOGS
AS ANTI-ANGIOGENIC AGENTS

By

Hari Srinivas Kalyan Paila

A Thesis Submitted to
the Faculty of the Graduate school at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Greensboro

2008

Approved by

Committee Chair

APPROVAL PAGE

This thesis has been approved by the following committee of the
Faculty of The Graduate School at the University of North Carolina at
Greensboro.

Committee Chair_____

Committee Members_____

Date of Acceptance by Committee

Date of Final Oral Examination

TABLE OF CONTENTS

	Page
GLOSSARY OF TERMS	iv
CHAPTER	
I. BACKGROUND INFORMATION - CANCER	1
II. TUMOR ANGIOGENESIS AND CANCER.....	9
III. THERAPEUTIC PROPERTIES OF CURCUMIN	12
IV. QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIP (QSAR) STUDIES.....	17
V. DESCRIPTORS IN QSAR STUDIES	24
VI. HYPOTHESIS AND EXPERIMENTAL DESIGN	36
VII. MOLECULAR MODELING AND COMPUTATIONAL DETAILS	41
VIII. DEPENDENT VARIABLE TRANSFORMATION	64
IX. RESULTS AND DISCUSSIONS	74
X. FUTURE RESEARCH	87
BIBLIOGRAPHY	88

GLOSSARY OF TERMS

Metastasis: The migration of cancer or tumor cells from the primary site of origin to random locations of the body via the blood or the lymphatic tissue to produce a secondary cancer growth.

VEGF: vascular endothelial growth factor.

bFGF: basic fibroblast growth factor.

MMPs: matrix metalloproteinases.

RI-QSAR: Receptor independent quantitative structural activity relationship study.

RD-QSAR: Receptor dependent quantitative structural activity relationship study.

Pharmacophore: Pharmacophore is defined as the three dimensional models of essential structural features of congeneric molecules that are necessary for biological activity.

Descriptors: Descriptors are the independent variables in a QSAR study. They provide information of molecules that can be correlated to their biological affinity in a QSAR or QSPR (Quantitative structural property relationship) study.

Force fields: Force fields are the energy functions of a microscopic system of particles (molecules). They indicate the total potential energy possessed by the specific system of atoms or molecules.

Dataset: Data set is defined as the biological affinity data of a congeneric series used in developing a quantitative structure activity relationship studies.

Training set: Training set is defined as the set of compounds in a dataset used to develop a QSAR model.

Test set (Predictive set): Test sets are the compounds used to externally validate the predictive property of a developed QSAR model.

Dependent variables: These are the biological affinity or biological response produced by the compounds of the dataset.

Independent variables: These are the molecular properties of the compounds. These are also called as descriptors in QSAR analysis.

Biological affinity: It is the biological response produced by the compounds in the assay.

Predicted values: These are the biological activity predictions of the QSAR model.

Actual (Experimental) values: These are the experimentally reported biological affinity data used in developing the structure biological correlation analysis.

Residuals: Residuals are the difference between the experimental values and the predicted biological activity values in a QSAR model.

Biological Affinity - ED₅₀: The amount of compound (concentration) that can produce a response in 50 % of the population.

LD₅₀ : The concentration of the compound that can produce a lethal effect in 50% of the population

Ki: The concentration of the compound that inhibits the enzyme completely in an enzyme catalyzed reaction.

ADME properties: Absorption, distribution, metabolism, and excretion are the pharmacokinetic parameters of drug like compounds.

CoMFA: Comparative molecular field analysis.

COMSIA: Comparative molecular similarity index analysis.

MSS: Molecular spread sheet.

PCA: Principal component analysis.

PLS: Partial least squares regression.

SAMPLS: Samples distance partial least squares; Regression analysis.

P: Inhibition potential (LD₅₀, ED₅₀, Ki)

CHAPTER I

BACKGROUND INFORMATION - CANCER

Among the subspecialties of internal medicine, medical oncology is an important frontier which has gained attention from various research communities. Cancer and its associated ailments are second only to cardiovascular diseases that lead to death of an individual. Demographical evidence shows one in every six deaths are caused by some type of cancer.¹ The mortality rate of cancer patients is usually high (65%). Appropriate therapeutic approaches are crucial factors in cancer treatment. Therapeutic approaches undertaken depend on the type of cancer, its pathological stage, and the condition of the patient. Primarily there are three treatment methods for cancer:² (a) Invasive methods e.g., surgical removal of cancer tissue (surgical oncology); (b) Irradiation methods, which involve the use of high energy radiation, (c) chemotherapy, which involve the administration of toxic drugs. Surgical methods are the most effective approaches since the bulk of the tumor tissue can be removed. Surgical removal of a mass is an option applicable for solid tumors and also for tumors that have not yet

metastasized. Metastasis is defined as the migration of the cancer cells from the primary region of origin to other parts of the body via blood and the lymphatic tissue. Usually, additional surgery is not an option if the tumor is in the vital organs such as the heart, lungs, nervous tissue, and kidneys.

Radiation therapy is an effective method as it can selectively destroy the tumor cells with minimal effects with surrounding tissue. Radiation therapy kills cancer cells directly or indirectly by inducing apoptosis (programmed cell death). Surgical and radiation therapy are the effective methods for a complete “cure” if the tumor is still benign and has not yet transformed to a metastatic form. Radiation therapy uses high energy radiation; usually the source is a radioactive decaying isotope.

Chemotherapy is the most widely used and effective method for the treatment of cancer, as metastasis is not a limiting factor in chemotherapy. There are various classes of drugs that target cancer cell growth and division. These chemotherapeutic agents exert their therapeutic action by hampering the nucleic acid metabolism of the cell directly or indirectly. These include (a) drugs that inhibit the synthesis of precursors for nucleic acid synthesis (antimetabolites); (b) drugs that

interfere with the replication of DNA (alkylating agents and DNA intercalating agents); (c) drugs that inhibit various enzymes in nucleic acid metabolism; (e) drugs that inhibits the formation of mitotic spindles during cell division.

Alkylating agents

These are the most widely used chemotherapeutic agents that were used to treat cancer in the 1960's and 1970's. They undergo activation to form strong electrophiles³. These electrophilic intermediates are attacked by nucleophilic groups such as hydroxyl, sulfhydryl, phosphate, amino, carboxyl, and imidazole groups of nucleic acids. The chemotherapeutic effects of these are due to alkylation of DNA. Alkylation occurs at the nucleophilic center of guanine (7th nitrogen position in guanine is the most nucleophilic atom) of the DNA strand. These alkylating agents are either monofunctional or bifunctional. In addition to the guanine, other nitrogens in purine and pyrimidine bases of DNA are also involved as nucleophiles. The 1st and 3rd nitrogen of adenine, 3rd nitrogen of cytosine, and the 6th oxygen of guanine are the nucleophilic centers.

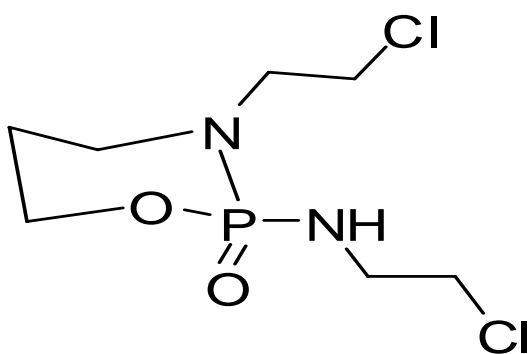
Nitrogen mustards were the first alkylating agents used in the early days of cancer chemotherapy. The below mentioned are some of the

widely used nitrogen mustards: mechlorethamine, cyclophosphamide, ifosfamide, melphalan and chlorambucil. The other classes of alkylating agents are dacarbazine, methylnitrosoureas (MNU): streptozocin, carmustine, lomustine, semustine and chlorozotocin. some are prodrugs, e.g. cyclophosphamide.

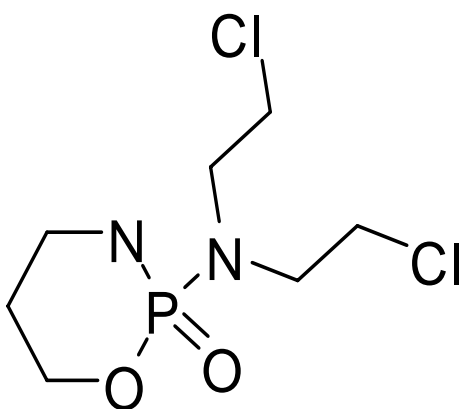
Antimetabolites-folic acid analogs

Folic acid is a precursor for nucleotide synthesis. Antimetabolites are a class of antineoplastic agents that inhibit the enzyme that catalyze the biosynthesis of nucleotides. These are the building blocks of nucleic acids, so inhibition of nucleic acid synthesis can retard cell division. Antimetabolites have structural resemblance with folic acid; they competitively inhibit the enzyme dihydrofolate reductase, which is a pivotal enzyme in nucleic acid metabolism.

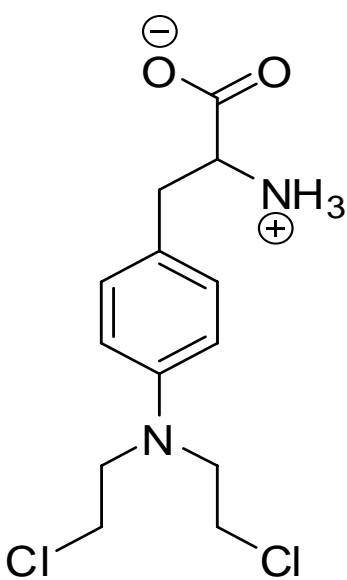
The structures of the aforementioned agents are some of the most widely used anticancer drugs that act by the mechanism of DNA alkylation.



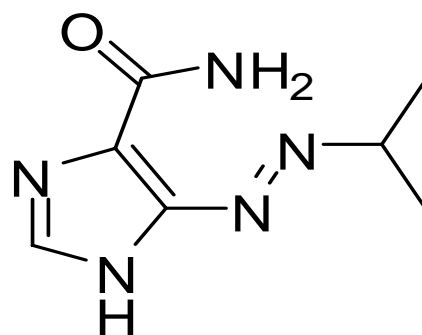
Iforsafamide



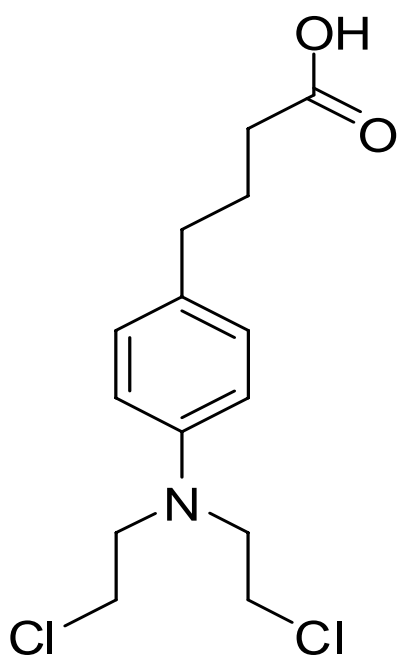
cyclophosphamide



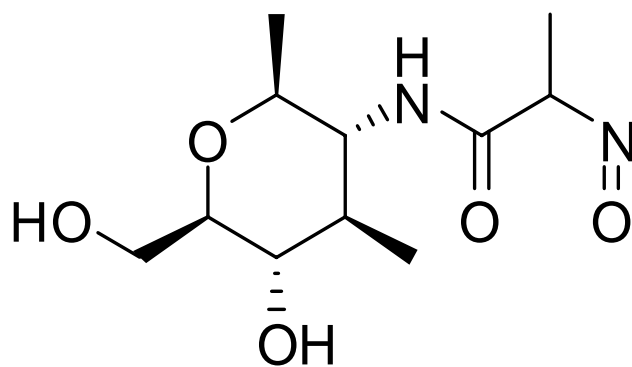
Melphalan



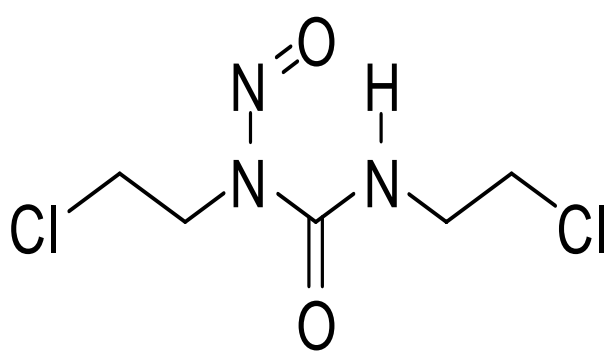
Decarbazine



Clorambucil



Streptozocin



Carmustine

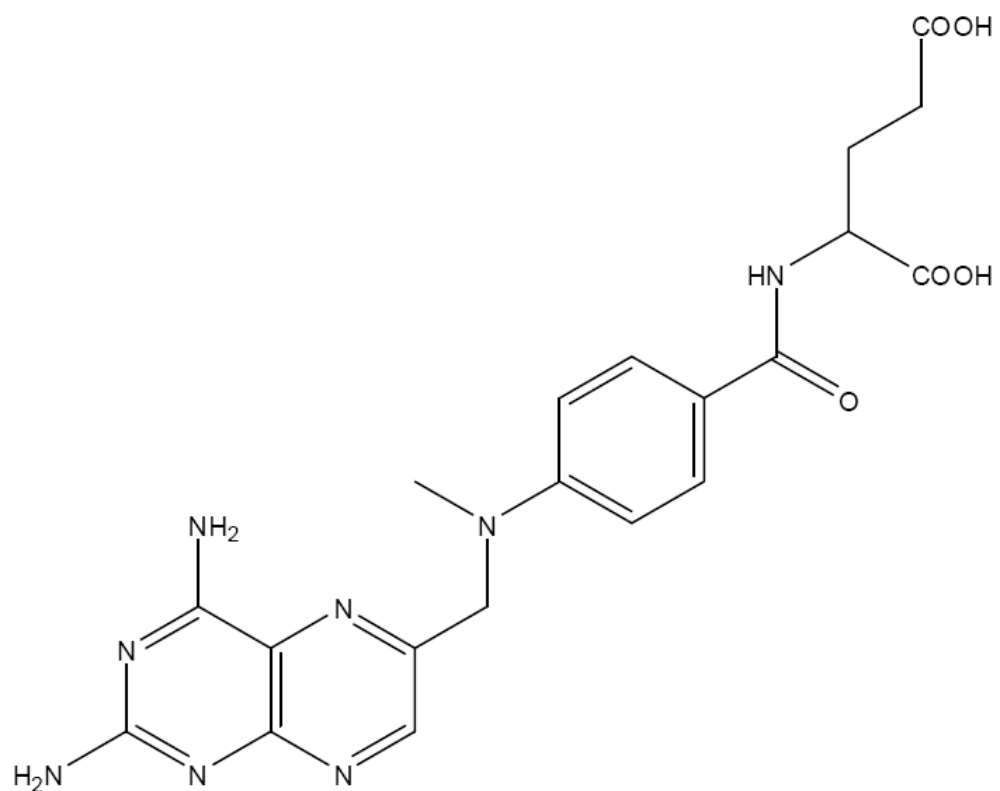


Fig 1.1. Antimetabolite (folate analogue) methotrexate. The compound has a structural resemblance to PABA (*para* amino benzoic acid), which is a natural precursor of folic acid and eventually depletes nucleotides in the cell.

Miscellaneous anti-cancer agents

There are miscellaneous drugs that act at various stages of the cell cycle. They include antibiotics, DNA intercalating agents, and drugs that inhibit the formation of mitotic spindles formation during cell division. The drugs in this category are taxol and *cis*-platin.

CHAPTER II

TUMOR ANGIOGENESIS AND CANCER

It is imperative for living cells to obtain nutrients, oxygen, and essential metabolites for their normal sustenance and growth. The diffusion limit for cells to acquire any of these nutrients is 100-200 micrometers, so any further beyond this point requires a direct blood supply.⁴ To render these vital necessities, cells recruit new blood vessels in the form of rudimentary endothelial cell based capillaries. This phenomenon is termed as vasculogenesis and angiogenesis. Angiogenesis is defined as the development of new rudimentary blood vessels in the form of capillaries from pre-existing vasculature. Angiogenesis is a phenomenon normally observed in several disease states like arthritis, corneal ulceration, proliferative retinopathy⁵, and tumors.

It is a well known fact that cancer cells divide at a faster rate and possess a higher metabolic rate than normal cells. Cancer cells promote angiogenesis via numerous biochemical messengers. Tumors can grow up to a certain mass (few microns in diameter) without direct vascular supply, but beyond this critical mass, tumors require direct blood vessels

for their survival and reproduction. This forces the tumor directly to stimulate angiogenesis from preexisting surrounding blood vessels.

Angiogenesis is observed as a normal phenomenon only in the physiological cases of embryonic development and morphogenesis, during wound healing, and in tissue regeneration. Pathological and pathophysiological angiogenesis is a hallmark of various ailments and diseases (eg. stroke, ischemic heart diseases, and rheumatoid arthritis). As evident from the above information, angiogenesis is a predetermining phenomenon for the survival and progression of a variety of solid tumors. J Folkman *et al* was the first to propose this hypothesis.⁵ Thus, therapeutic intervention of the multiple targets that promote angiogenesis can significantly retard or completely inhibit angiogenesis and eventually suppress cancer growth by indirectly cutting off the supply of various necessities for normal cell growth.

Angiogenesis is a complex phenomenon, and the biochemical mechanisms that control it are delicate in nature. Angiogenesis is controlled by a sensitive physiological switch, which is triggered either by proangiogenic factors or antiangiogenic factors. These angiogenic factors are diverse in their function and broadly distributed both in

intracellular and extracellular environments. In normal tissues antiangiogenic factors predominant over proangiogenic factors. The most prominent proangiogenic factors are VEGF (vascular endothelial growth factor), bFGF (basic fibroblast growth factor), MMPs (Matrix metalloproteinases), cyclo-oxygenase - 2 (COX-2).⁶ the antiangiogenic factors are VEGFR-1, VEGFR-2, vasostatin, and endostatin.

CHAPTER III

THERAPEUTIC PROPERTIES OF CURCUMIN

Curcuma longa L is a tropical herb usually seen in Southeast Asia. Since medieval times, the herb was reported in the literature for its various therapeutic effects. The curative property was attributed to the yellow pigment constituent in the rhizome of the plant. This active principle is curcumin (diferuloyl methane). Curcumin has a broad range of therapeutic properties. In recent times *in vitro* studies of curcumin and its various synthetic analogues demonstrated several and diverse categories of biological activity, especially anti-inflammatory (Crohn's disease, arthritis, and several disorder of the cardiovascular system) and antimetastatic-antiangiogenic (anticancer potential and several other miscellaneous diseases). In addition to the above, it also has anti-oxidant, anti-viral, and anti-infective activity. Curcumin with its antioxidant and anti-inflammatory activity has invaluable therapeutic advantages. Experimental studies have demonstrated that curcumin is a potent scavenger of reactive oxygen species, which include superoxide anion radical and hydroxyl radical.⁷ Various pathophysiological effects of curcumin at the cellular level include: induction of apoptosis

(programmed cell death) in cancer cell lines through downregulation of proangiogenic genes mediated by transcription factor NF- κ B⁸ and IKB kinase, consequently. Arrest cancer cell growth in S, G₂ and M phases of the cell cycle.⁸ Northern blot analysis indicates a time dependent down regulation of VEGF and angiopoietin; these two targets are the most potent stimulators of angiogenesis. With the above mentioned basis of anticancer potential and potent antiangiogenic property against various biological targets. Curcumin and several synthetic analogs of curcumin were studied extensively. Inhibition of arachidonic acid metabolism is the most prominent feature of curcumin in cellular system. Arachidonic acid is also a key promoter of carcinogenesis in living systems.

A significant role of curcumin is its suppression of metastasis and the progression of solid tumors. Curcumin has been shown to inhibit the differentiation of human umbilical vein endothelial cells (HUVEC).⁹ It also inhibits bFGF - induced corneal neovascularization in the mouse corneal. This phenomenon was discovered by Arbiser *et al.*, and curcumin was considered as a promising lead candidate in biomedical research. So, several curcumin analogs were synthesized and tested for anticancer and angiostatic activity by various research groups using several bioassays with different cell lines.

Bioactivity studies have shown that various curcumin analogs are angiogenesis inhibitors, as observed in the chorioallantoic membrane assay. These inhibitory activities of curcumin analogues are due to reduction in the activity of vascular endothelial growth factor (VEGF) and matrix metalloproteinases (MMPs). Matrix metalloproteinases play a major role in cell membrane restructuring and consequently facilitate angiogenesis.

There are several biochemical mediators that promote cancer progression. The biological activity of curcumin can be attributed to the ways it affects various modulators of cellular and subcellular receptors (EGFR and HER2), transcription factors (NF- κ B, AP-1, Egr-1, beta catenin, and PPAR- γ), and various inflammatory mediating enzymes like cyclooxygenase (COX-2), 5-lipoxygenase, nitric oxidase synthase (iNOS), cyclin protein D1p, and cytokines (TNF, IL-6, IL-1 and chemokines). The biological activity of curcumin was attributed to the electrophilic property of the β diketone linker.¹⁰

With these diverse bioactivity profiles, curcumin is considered a potential natural product lead for structure optimization in drug

discovery. But its poor solubility, undesired pharmacokinetic-pharmacodynamic profile, and the lack of selectivity towards a single target limits its utility. Structural modifications of curcumin analogs are an available alternative. Novel analogs may be designed through *in silico* methods, synthesized, and subsequently tested for their biological affinity towards a single biological target of interest. Curcumin is a benchmark reference.

Molecular features of curcumin

Curcumin exists in keto-enol tautomeric form, characteristic of 1,3 carbonyl systems and there is intramolecular hydrogen bonding present in the molecule. The herbal extract of *Curcuma longa* has three fractions: curcumin (77%), demethoxycurcumin (17%), and Bisdemethoxycurcumin (3%). The latter two are collectively called curcuminoids. The central β -diketone linker is an essential pharmacophoric feature for biological activity, and the two aromatic groups attached to the linker can be structurally modified for desired pharmacokinetic factors and selective biological affinity.

Structure of curcumin

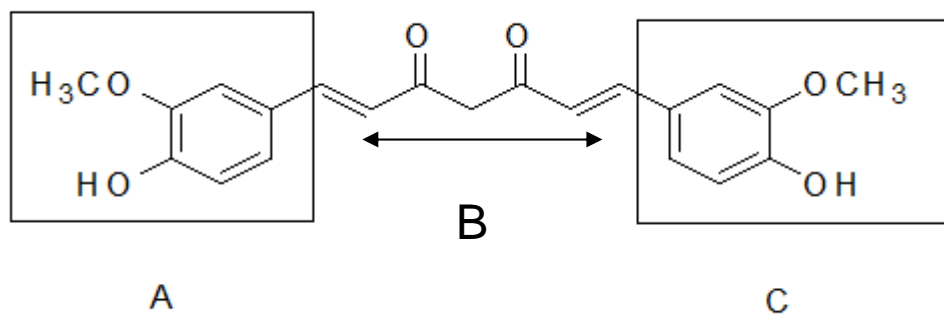


Fig 3.1. Three regions of curcumin: A and C are regions accommodating aromatic systems, and region B is the diketone linker.

The following are the therapeutic activities of curcumin/turmeric; Antioxidant activity, anti-inflammatory, enhance wound healing, immunomodulatory, antispasmodic activity, antifungal, antiparasitic and antibacterial activity, antimutagenic, antimetastatic, and antiangiogenic activity.

CHAPTER IV

QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIP (QSAR) STUDIES

In silico molecular modeling methods are extremely useful in the biomedical sciences. Fundamentally, molecular modeling methods are classified into structure - based or structure - dependent and ligand - based or structure - independent methods. Here, structure refers to the biological target (receptor, enzyme, nucleic acid, ion channel, or carrier protein). QSAR analysis, pharmacophore modeling, putative receptor modeling, and consecutive search query fall in the latter class of molecular modeling methods. The predictions of these ligand based methods are pivotal in making critical decisions in future analog synthesis.

The primary intent of this thesis is to develop a three - dimensional QSAR model of experimentally validated compounds (curcumin analogs) as anti-angiogenesis agents. The QSAR model development and its predictions are used to develop synthetic targets of potentially active compounds. Presumably, this will significantly reduce research time and expenses. A QSAR, in general, may be a reliable tool when structural

information of the biological target is unknown – i.e., has not been characterized by X - ray crystallography or by NMR methods. When the drug - receptor structural details are available however, it can yield more realistic picture which in turn, leads to robust computational models with higher probability for activity prediction.

Microscopic characteristics of matter are important for compounds possessing biological activity. In biological systems, these properties are the results of the constituent atomic composition and arrangements within the molecules. Most biomolecules, natural ligands, and drug-like molecules usually interact with their biological target through non covalent molecular forces. These Interactions are usually shape dependent. Non-covalent interactions fundamentally depend on the magnitude of steric and electrostatic forces. Computational drug designing strategies are fundamentally classified into two methods: (a) structure (receptor)-based molecular modeling methods, and (b) ligand - based molecular modeling methods. When the biological structural data (receptors, ion channels, enzymes, or Nucleic acids) are available, transition state analog design, docking studies or molecular dynamic simulation studies can provide an in-depth picture of the interaction of the active ligand with the interacting elements in the active domain of the biological target. The prerequisite for drug-like molecules is to have high

affinity at the active site of the biological target when compared to the natural ligand. It has been postulated that the pharmacodynamic activity is achieved through drug-receptor binding with subsequent conformational changes in the drug receptor complex, which results in a cascade of biological events leading to the observed physiological change. This is achieved when the affinity for a drug like molecule is higher than the natural ligand in a biological system.

When the biological structure data are unknown, structure-activity correlation studies with statistical methods can provide information regarding the characteristics of the putative target/receptor. Structure-activity relationship studies (SAR), *de novo* ligand design, and semi-empirical calculation methods are some of the other alternative approaches used in ligand design. Several computational methods are used to develop structure-property relationships (QSPR). QSAR empirical studies are one of the extensive methods used to predict structure-property co-relations of bioactive molecules. In the drug discovery process various properties (e.g. physicochemical, quantitative structure property relationship (QSPR), pharmacokinetic and toxicological) are important. Quantitative structure toxicity relationship (QSTR) are to be evaluated collectively with an intention to develop a successful clinical candidate. A QSAR study can be undertaken to

predict any of the above mentioned properties. These structure-property correlation studies are pivotal in the drug discovery pipeline. QSAR studies provide an invaluable tool for identifying the important structural features required for activity, and eventually a computational tool for lead optimization.

Minor changes in the constituents of the molecules, functional groups changes, or alternation in geometry can affect biological activity: stereo chemical properties in a molecule for a series of structural congeners can profoundly affect biological activity. Three-dimensional quantitative structure activity relationship (QSAR) analyses can provide an in-depth analysis based on two most likely interaction forces: steric and electrostatic.

When biological activity for a series of molecules which differ in their structural features is observed, empirical methods like docking, scoring, and structure property or structure activity relationships can be used. QSAR in a pharmacodynamic perspective is a statistical correlation method between molecular properties and their corresponding biological activity *in vitro* or *in vivo*. This approach has the potential to predict for a wide range of properties (e.g., water solubility, lipophilicity, partition function and biological affinity). Also, QSAR aids in differentiating

drug-like molecule from non drug-like molecules based on toxicity predictions, drug metabolism, and ADME properties, as well as drug deposition in the tissue.

The vast majority of target specific drugs fall in either one of the following categories: organic small molecules, nucleotides, oligosaccharides, peptides, or proteins. The most significant properties of these are steric (shape and volume) and electronic (electrostatic potential and electric charge). The above mentioned two are the major properties of drug-like molecules for their affinity towards their biological target.

A dataset of a homologous series of molecules (target specific ligand or molecules) that has biological affinity towards a specific target can be modeled to predict the properties of unknown molecules. QSAR studies were first developed by Hanch and Fujita.¹¹ The proposed study was based on substitution constants. These quantitative structural-activity relationship studies are based on how the property is being affected as a function of the substitution groups in a series of congeners. These substituent constants are hydrophobic constants (π), Hammett constants (σ), and molar refractivity (MR). The hydrophobicity constant (π) provides information of the logarithm of the partition constant ($\log P$) of a substituted molecule relative to an unsubstituted molecule. The

hydrophobicity constant is an indicator of the non-polar/polar nature of the substituent. The Hammett constant σ is an indicator of electron-donating or electron-withdrawing effects of the substituent. It is derived from the ionization constant of acids, and the substituent present in the acid molecule. Molar refractivity is an indicator of polarisability of the substituent on the molecule, and it is derived from the Lorentz-Lorentz equation. This equation is a function of density and molecular weight of the compound and its refractive index. QSAR studies can be feasible with the above methods when there is a similar motif with the principle structure and few substituents. When there is structural diversity and increased three dimensional conformations. QSAR analysis with the above mentioned constants become a difficult task to try to correlate structure with bioactivity. Even though correlation can be achieved, validation of the results is a difficult task. Later the two-dimensional QSAR methods were developed, which are based on the topographical indices, but these methods were also found to have some drawbacks.

Three-dimensional QSAR methods are becoming the most extensively used QSAR methods in property prediction in biological systems, as well as in toxicological studies. As mentioned before, the interaction is dependent on the three-dimension arrangement of groups in the molecule. The independent variables used in QSAR studies are the

indicators of the microscopic properties of molecules. These can either be derived or calculated by quantum mechanical methods. These independent molecular information rich variables are termed as descriptors. Descriptors provide information of the molecules and how it affects their physicochemical properties.

CHAPTER V

DESCRIPTORS IN QSAR STUDIES

Descriptors are information rich variables defining molecules. Descriptors are grouped in two categories: (a) whole molecules descriptors, and (b) fragment descriptors. Whole molecule descriptors are indicator properties of the entire molecule itself (e.g., molecular weight, molar refractivity). Fragment descriptors are calculated based on the constituent atoms or substituent functional groups of the molecule. Substituent constants like the hydrophobicity constant (π) and the Hammett constant (σ) fall in this category.

Molecular descriptors can be calculated by computational methods. Currently, molecular descriptors are calculated readily with currently available computer hardware and computational software. Fragment based descriptors are independent of the configuration and the three-dimensional confirmation of the molecules. They are not as information rich when compared to whole molecule descriptors. Whole molecule descriptors are efficient and widely used descriptors, as they provide information of three- dimensional properties of molecules.

The classic whole molecule descriptor is clogP, which is the octanol-water partition function. Partition of drug molecules through the biological membrane into the cell is an important factor to exert its biological activity. Other three-dimensional descriptors are derived from molecular fields i.e., steric and electrostatic fields.

Judicious selection of descriptors or a group of descriptors is critical for the intended QSAR analysis and property prediction. Descriptor selection is based on the information content and its degree of relevance to the QSAR study. Some descriptors provide desired results for one QSAR property prediction, while other types of descriptors are useful for different types of property predictions. There is some interdependency between descriptors, so when they are used as independent variables in a QSAR study they yield similar results. Thus, an appropriate selection of descriptors is an important factor. Poor selection of descriptors yields inefficient QSAR models or even gives misleading predictions. Selections of many descriptors are to be avoided as they might tend to over fit the model. This can be due to chance correlations because many descriptors were tried in developing the model. Chance correlation is a major setback in QSAR model development. To avoid chance correlation, variable selection methods

through genetic algorithms and progressive scrambling are carried out. This can avoid the potential problem to some extent. But the modeler is to be always wary of these factors during QSAR model building.

QSAR method based on Comparative Molecular Field Analysis – CoMFA and Comparative Molecular Similarity Indices Analysis – CoMSIA

Biomolecules interact significantly through steric and electrostatic forces, so it is obvious to use descriptors derived from these two contributing fields in the QSAR model development. Based on this hypothesis, comparative field analysis (CoMFA) and comparative molecular similarity indices analysis (COMSIA) QSAR methods were developed. First, CoMFA analysis was developed for a series of steroid and their binding affinity towards carrier proteins by Richard D. Cramer *et al.*¹² This work laid the foundation for computational studies in which shape (three-dimension arrangement of atom or groups in a molecule) can significantly affect their affinity to a biological target. Three-dimensional QSAR studies based on the CoMFA method or alternative methods are considered to be reliable for property predictions of unknown compounds.

There exists a relationship between molecular structure and biological affinity: BA (eg., K_i , IC_{50} , LD_{50} , ED_{50}). This biological response is expressed as $\text{Log} (1/BA) = f (x_1, x_2, x_3, \dots, x_N)$. The descriptor variables, $x_1, x_2, x_3, \dots, x_N$ (steric and electrostatic fields) are determined by PLS (Partial least squares) analysis.

Three-dimensional QSAR methods based on CoMFA have some inherent inconsistencies that can affect the final predictive property of the model. They were addressed later by Gerhard Klebe *et al.*¹³ His research group implemented similarity indices as descriptors in three-dimensional QSAR model development. This formalism of the QSAR model development is termed as comparative molecular similarity analysis (CoMSIA). This method is more successful than its predecessor CoMFA based QSAR models. The important drawback of CoMFA is its steep region shift; the magnitude of Lennard-Jones and Coulomb potentials change is high and increases as the sampling is performed from the outer regions towards the center of the molecule. To avoid this steeper region shift in the QSAR model development, cut-off values are used for Lennard-Jones potential (steric fields) and the Coulomb potential (electrostatic fields) calculation in the CoMFA based QSAR model. Actually the regions inside the molecules have some information rich

descriptors. Potential $E(r)$ is inversely proportional to the distance (r) from the center of the molecule. CoMFA used Lennard-Jones and Columbic potentials in the calculation of the descriptors while CoMSIA uses Gaussian approximation in calculating similarity indices at various grid points.

The protocol for developing a three-dimensional QSAR model is similar in both CoMFA and CoMSIA.

Three-dimensional QSAR analysis based on CoMSIA and its advantages over CoMFA

CoMSIA analysis uses sampling of similarity indices at various grid points of the common alignment of molecules in the dataset. This sampling is performed at all the equally spaced grid points of the defined lattice. These lattice points are present inside and outside the molecule. In CoMFA analysis the lattice points located inside the molecules are not taken into account in the statistical correlation analysis. For example in PLS analysis¹⁵ there is a cut-off value defined before the PLS run is carried out. This cut-off value is assigned to avoid field values that are very high in magnitude. But in a CoMSIA analysis, the lattice points

present inside and outside the molecule are taken into account. The potentials at the lattice points are the independent variables used in PLS analysis. CoMSIA has five different property fields. These five different fields are used to significantly evaluate the property and partition them spatially in different locations of the molecule. This, in turn, signifies which property plays a critical role in their contribution to the affinity or binding of the molecule to the receptor. It has significant importance in several steps of lead optimization when diverse properties of drug molecules are taken into account.

Designing the calibration set or training set

In any QSAR method the fundamental basis for developing the model is universal. There should be a statistical correlation between the dependent variable and the independent variables (descriptors). These relationships are either linear or nonlinear in nature. The dependent variable is usually the biological affinity (K_i , IC_{50} , LD_{50} and ED_{50}), which is converted to Log_{10} (1/activity): pED and pK_i . A QSAR analysis is usually carried out in two steps.

The first step is to develop a model that describes the correlation between the independent and dependent variables, which

is defined as this function: $y = f(x)$. The relationship is expressed as a function, and this defines how the dependent variable (y) is affected by the independent variables (x), the molecular descriptors. The training set model parameters are called regression coefficients or sensitivities. The later step is to derive independent variables from one or more samples and the sensitivities from the model.

In step two the derived independent variables from the training set are used to predict the dependent variables (biological affinity). This prediction is performed with the test set or the prediction set. This is the reason why the initially chosen dataset is divided into two groups: calibration set (or training set) and prediction set (or the test set).

The training set is used to develop the QSAR model, and the test set is used to validate the model externally and also to predict the degree of accuracy of the model. The robustness of the model depends on the predictable properties of the model. The dataset should be large enough with the diversity in the structural congeners and their corresponding range of biological activity to be meaningful. The training set compounds are usually higher in number than the test set. The dividing of the training set and the test set is based on random picking of the compounds and

classifying them into the training set and the predictive or test set. The selection of training set and test set is done taking into account the structural diversity in both the datasets and the corresponding range of biological activity of the molecules.

Three-dimensional QSAR method based on CoMFA and CoMSIA analysis are performed with the following protocol

- 1) Structural input via sketching the molecule in the computer program followed by energy minimization, and calculation of mathematical descriptor is carried out. The later is the essential property of the molecule in three-dimensions. One should take into account which descriptor is efficient in structure property correlation studies.
- 2) The molecules of the dataset are aligned so that they have a common three-dimensional orientation. Alignment is usually carried with atom fit or field fit alignment methods in Sybyl. Several modifications can be carried out (changing the stereochemistry, modifying the torsion in the common backbone). Alignment is carried out in a manner to yield the least possible

(RMS) root mean square deviation between the common points chosen in aligning.

- 3) An evenly spaced (step size) lattice on the superimposed alignment is defined. The dimensions of the grid are with a little extension in all the three x, y, z - cartesian co-ordinates so that it accommodates all the molecules. Usually the grid spacing of the lattice is 2 Å; however, the grid dimensions and the spacing can be modified if the superimposed molecules are to be rotated or translated in the grid. The step size or spacing of the lattice can be modified to 1.5 Å or 1 Å. This is observed with a penalty of increased computational time in developing the model as the number of PLS sampling points increase exponentially in the lattice.
- 4) The magnitude of the steric (van der Waals) and electrostatic fields (Coulombic with a distance dependent dielectric- $1/r^2$) at all the lattice points are calculated. The steric and electrostatic interactions are determined with a probe atom (sp^3 hybridized carbon atom with a unit positive charge).
- 5) A molecular spread sheet is created with the dataset. Converting the dependent variable, biological affinity (IC_{50} , LD_{50} , ED_{50} , Ki)

to Inverse logarithm values (1/biological activity) and entering the data as the dependent variable in the QSAR analysis is required.

- 6) The columns are created with the independent variable (CoMFA columns and CoMSIA Indices as per the analysis whether it is CoMFA or CoMSIA).
- 7) Statistical correlation analysis is carried out in a sequential manner with leave-one-out PLS method first. This is done to identify the optimum number of components to derive the best model. Then these optimum number of components are used in cross-validation to predict the q^2 (average r^2). The cross validated correlation coefficient q^2 falls in the range from 0 to 1. For really good predictable models, the cross validated q^2 is in the range of approximately 0.4 and higher. The final step is the no-validation PLS method.
- 8) The last step is simultaneously predicting the dependent variables of both the training set and the test set.
- 9) Residuals or errors for the compounds in the QSAR model (training set and test set compounds) are determined. Residual or errors of prediction are calculated as follows: Residuals = experimental activity – predicted activity.

- 10) The scattered plots of predicted vs experimental or actual dependent variable (biological affinity) are generated. If the model is reliable, there is a positive correlation and most of the data points are pretty close to the 45° diagonal in the scattered plot; however, there usually will be some outliers in the model.
- 11) With the magnitude of residuals in the scattered plots, compounds that are highly predicted and outliers can be observed in the model.
- 12) After eliminating the outliers, the QSAR models are derived in an iterative manner to check the improvement of predictions of the model. Elimination of too many outliers can result in a propensity for a biased model.
- 13) The final step is the validation of the QSAR model. This is the external validation of the model with the compounds of the test set. The predictions of the QSAR model are interpreted as stereo contour coefficient maps. Compounds with good predictions by the model are placed in the contour coefficient maps. Now the favorable and disfavorable interactions of different field contributions, which are represented as contour maps are observed. These maps provide insights into any structural modifications in the different regions of the compounds (addition

or elimination of electrostatic - charged groups, steric - bulky groups, H - bond donor or H - bond acceptor) and may be inferred with stereo view of the contour maps.

CHAPTER IV

HYPOTHESIS AND EXPERIMENTAL DESIGN

The primary focus of this thesis is to develop a three-dimensional QSAR model based on CoMFA fields and COMSIA indices as descriptors. Since the structural information of the biological target is neither available in the SVR cell growth assay nor reported in literature elsewhere, ligand based modeling is the only method to develop a computational model for the prediction of potent curcumin analogs. This study should be able to pave the way for further research towards the synthesis of new molecules.

Based on whether there is structural information available or not, QSAR studies can be carried out by two methods: (1) It can be either structure independent or (2) structure dependent. For the former, the biological target structure is not considered directly in developing the QSAR model. The later approach, however, includes structural details of the biological target into account when developing the model. In this research study, the precise biological target on which curcumin acts is not certain; therefore, the structure independent QSAR methods were employed in developing the model.

Three-dimensional QSAR methods are most widely used to predict new chemical entities *in silico* which can possibly have improved biological activity. The Sybyl¹⁹ (versions 7.2 and 7.3) molecular modeling software was used as the modeling tools in developing the QSAR model (based on CoMFA fields and COMSIA indices as descriptors). The computational methods employed in this research are entirely structure independent methods. The dataset of curcumin analogs aromatic enones with their corresponding biological activity was determined via the SVR assay and previously reported.¹⁸

Details of experimental assay data: Dataset

The endothelial cell proliferation assay was developed Dr Jack Arbiser *et al* at *Harvard* medical school.¹⁶ Based on this experimental evidence of *in vitro* SVR cell growth assay of curcumin, curcumin's biological activity can be attributed to the inhibition of basic Fibroblast growth factor (bFGF) whose effects are shown to inhibit endothelial cell proliferation and angiogenesis *in vivo*.⁶ With this evidence, various structural analogues were explored for biological activity. Curcumin has three distinct molecular regions: the two aromatic systems (R₁ and R₂) and the central hepta-1,6-dien-3,5-dione linker. Several structural analogues of curcumin were designed and synthesized. The

corresponding SVR assay was carried out with all the aromatic enones. There is a significant increase in potency observed with these compounds. The inhibition potential ranged from 0 to 100 %. These aromatic enones have a central linker, this linker is abridged to an enone system instead of having hepta-1,6-dien-3,5-dione linker found in curcumin. This is followed by the substitution on the aromatic system with various substituent groups R_1 and R_2 or with heterocyclic aromatic systems. The aromatic enones curcumin analogs were designed and synthesized in the Bowen research laboratory at the University of Georgia. Some of the compounds may have been obtained commercially, if available at that time.

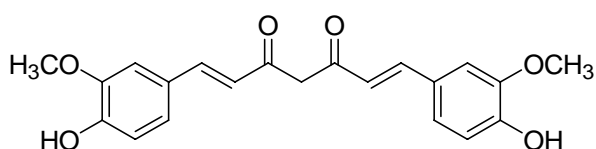


Fig 6.1. Curcumin

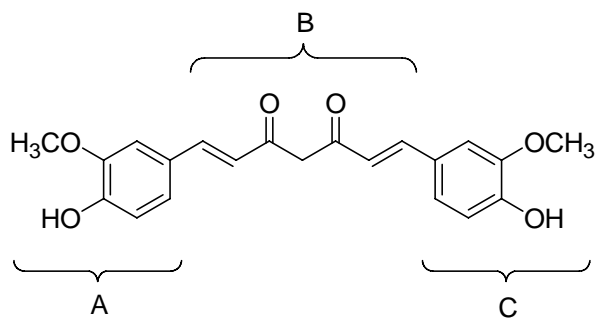


Fig 6.2. The three regions of curcumin A and C form the Aromatic systems and C the β - diketone linker.¹⁷

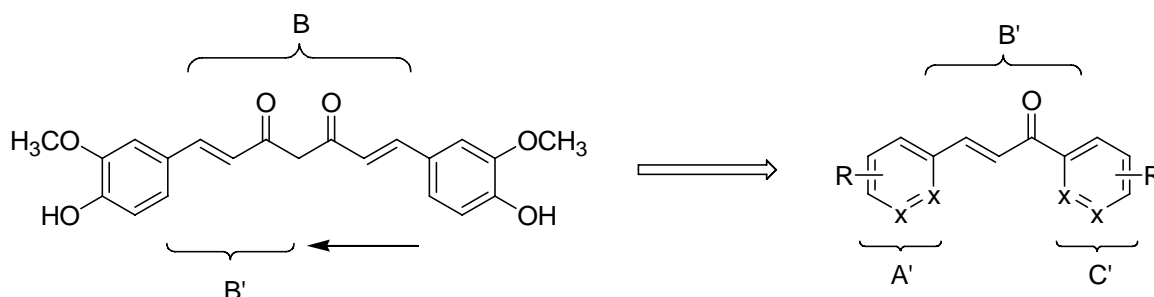


Fig 6.3. Curcumin structure and the abridged aromatic enone parent Analog. X==X represent various cyclic linkers.¹⁷ R represent various substituents on the two aromatic systems A' and C' respectively.

The *in vitro* SVR assays of the sixty three aromatic enone analogs were carried out at Emory University School of Medicine. The assay is a cell based assay with 10,000 murine cells/well in a 24 well cell culture dish. Initially the cells were incubated with 10 % DMEM (1ml/well) under CO₂ atmosphere. During the 24 hrs incubation period, the cells adhere to the bottom of the surface. After the 24 hrs period the media was replaced again with a 10 % DMEM media. Simultaneously all the sixty three compounds were dissolved in the least amount of dimethylsulfoxide, DMSO, to prepare a stock solution of 10 µg/ml, with appropriate dilution concentrations of 1 µg/ml, 3 µg/ml, 6 µg/ml, or 9 µg/ml. The cultures were later incubated for 48 hrs, and the cells were washed. The number of cells per well was determined with Beckman Coulter Z1 cell and particle counter, and the percentage inhibition was calculated. Finally this percentage inhibition was reported at three different concentrations of 1 µg/ml, 3µg/ml, and 6µg/ml for each compound.

CHAPTER VII

MOLECULAR MODELING AND COMPUTATIONAL DETAILS

Many of the aromatic enone antiangiogenic agents used in the dataset have been published.¹⁸ The dataset consists of sixty three compounds and their corresponding biological activity. The biological affinity as measured by percent inhibition ranges from 0 to 100% was experimentally determined for all the sixty three compounds at three different concentrations (1 μ g/ml, 3 μ g/ml and 6 μ g/ml). These compounds were evaluated in the SVR cell growth assay as discussed in chapter 6. The SVR assay was termed after the cell lines used in the assay. SVR is designated for Immortalized endothelial cell. In the development of the QSAR model, the dataset was grouped into calibration or training sets consists of 46 compounds, and the prediction or test set or predictive set of 11 compounds.

Structure sketching and alignment

All the structures were sketched with the cleanup options available in Sybyl 7.2 of Tripos Inc.¹⁹ Gasteiger-Huckel partial charges were calculated for all the molecules. Energy minimization was carried out

with the following minimization parameters to produces a local minima conformation of each molecule.

Force field: Tripos force field²⁰ was used in the energy minimization of every molecule.

Charge: Gasteiger-Huckel partial charges are calculated for all the molecules in the dataset. Distance dependent dielectric constant $(1/r) = 1$ is used in calculating the interaction fields at various lattice points. Conjugated gradient method was used in energy minimization with 5000 minimization iterations. Step size of 0.005 kcal/mole was set in the process.

The energy minimization was carried out with a 0.005 kcal/mole gradient step size with conjugated gradient as the termination method. The numbers of iterations used were 5000. This energy minimization yielded the local minima for the molecules. All the molecules were named as (Cur_1, Cur_2, Cur_3,.....Cur_63). A database named CurALIGN40.mdb was created with these compounds.

Database alignment

The next step is the alignment of molecules of the dataset: database. Alignment is achieved taking into criteria of the co-ordinates of the molecules that render the final common alignment with least root mean squares (RMS) deviation in the common groups or atoms used in aligning the cogenic series. Typically, the alignment of the compounds for QSAR studies is usually performed by three methods.²¹

1) Bioactive conformation based alignment (BCBA)

This method is followed when a known bioactive conformation of a compound is available as co-crystallized with the target structure.

2) Conformational search conformation based alignment (CCBA)

In this method conformational search is performed with to observe the global minima. Alignment is carried out with the global minima of all the compounds in the dataset.

3) Docked conformation based alignment (DCBA)

In this method the compounds are docked into the interacting domain of the target structure.

4) Local minima confirmation based alignment (LCBA)

Here the compounds are sketched and energy minimized to yield a local minima. The local minima conformation of each molecule in the database is used in the alignment procedure.

Rigid atom fit or field fit alignment are the most common types of alignment rules followed. Both these alignment are performed keeping in mind that there is a minimum RMS deviation in the common points or atoms chosen in aligning the database. The RMS deviation observed in the initial trial alignment of the dataset of the first forty two compounds was acceptable, but beyond compound forty (Cur_40) there is an observed significant root mean squares (RMS) deviation in the alignment of the molecules. The degree of predictive properties of a three-dimensional QSAR studies based on CoMFA and CoMSIA analysis are highly depends on extent of closeness in the common points chosen in aligning the database. During the execution of the energy minimization protocol of each molecule, the compounds attain local energy minima. During this process the molecules assume a structure with the local minima energy conformation. This local energy minimum varies from one molecule to the other in the dataset: It depends on different atom types and varying substituents of the molecule. In three-dimensional QSAR studies, the alignment is the most important aspect to be considered to yield a better crossvalidated q^2 (average r^2) and eventually the predictions of the computational model.

Thus, several methods and optimizations were carried out with the molecules to yield a best possible alignment. Dihedral angles were modified for the analogs to achieve the desired configuration for the compounds that were not aligned properly when superimposed. Two centroids (C_1 and C_2) were defined for the two aromatic substituents R_1 and R_2 . The centroids C_1 and C_2 and all the atoms labeled 1, 2, 3, 4, 5 and 6 of the central enone linker were chosen as common points in aligning the database. The below figure shows the points chosen for alignment of the structures.

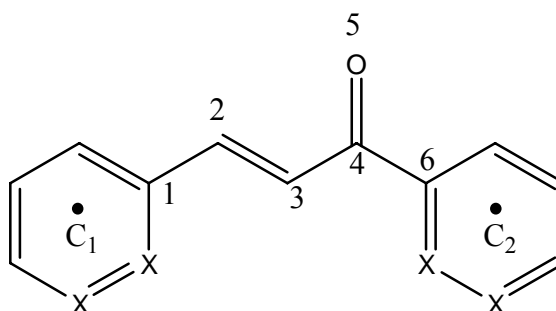


Fig 7.1 C_1 and C_2 centroids of the aromatic rings were done with the entire labeled atoms 1 through 6 in the central chalcone bridge. X represents various heteroatoms in the aromatic systems.

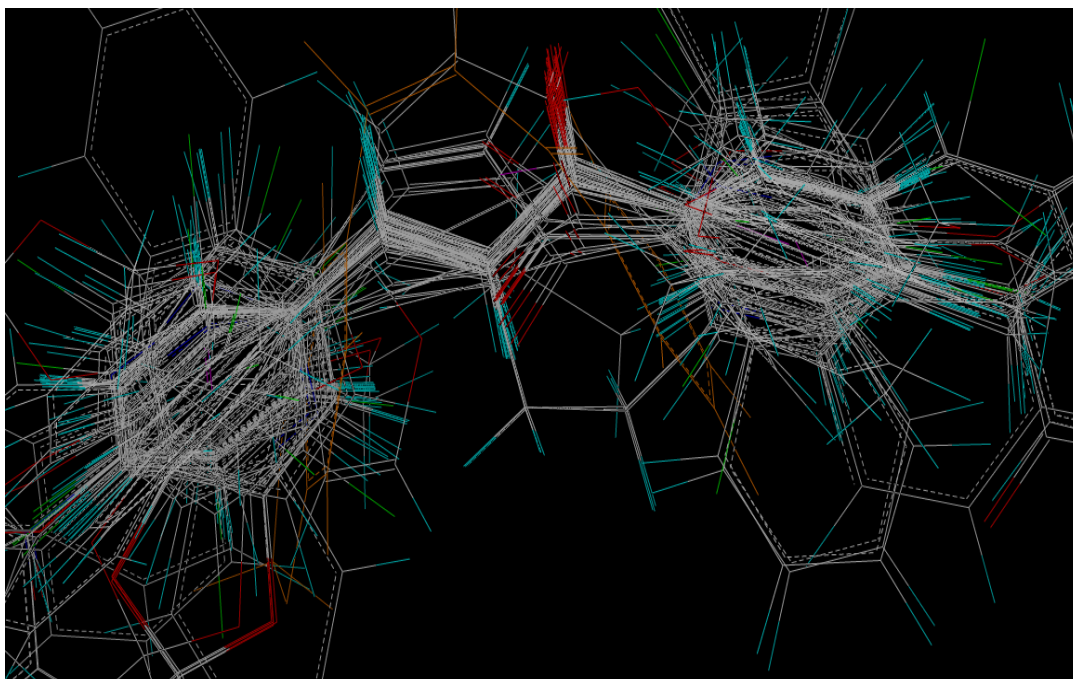


Fig 7.2 Initially developed trial alignment: poorly aligned regions are seen in the central enone linker.

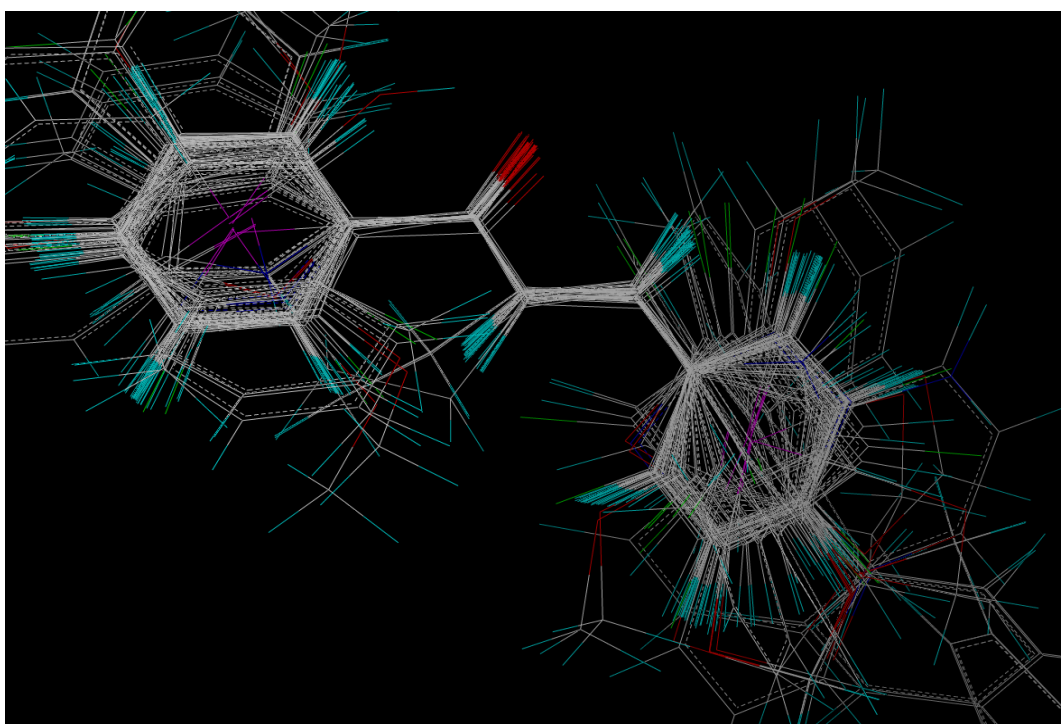


Fig 7.3 Rectified alignment: This alignment was achieved by atom fit of the central enone linker and the centroids C_1 and C_2 defined for Aromatic systems R_1 and R_2 .

Statistical analysis of the model: PLS (Partial Least Squares)

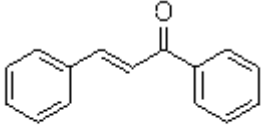
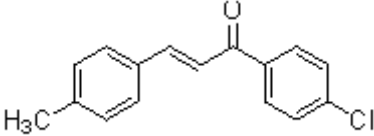
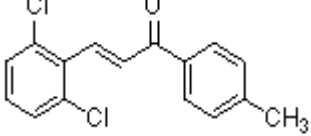
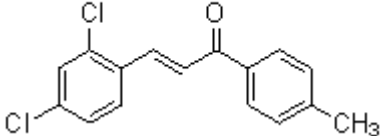
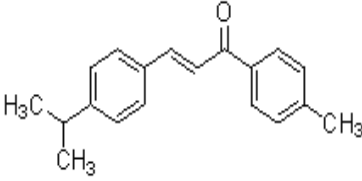
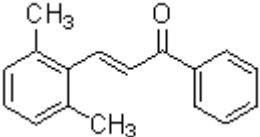
PLS (partial least squares) analysis is a statistical correlation analysis.²⁹ PLS analysis is used with this dataset. The natural logarithm of percentage inhibition, \ln (% inhibition) is used as a dependent variable. The range of the dependent variable for the dataset is observed to be in a range of 5 log units. Steric and electrostatic field descriptors are used as the independent variable in three-dimensional QSAR CoMFA model development. Similarly, five indices; electrostatic, steric, hydrogen bond donor, hydrogen bond acceptor and hydrophobicity indices are used as dependent variable in three-dimensional QSAR model development based on CoMSIA.

PLS analysis with CoMFA descriptors

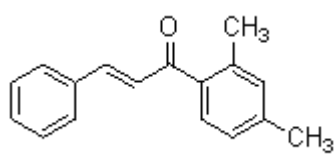
The dataset of sixty three compounds (Cur_1, Cur_2, up to Cur_63) was used in the analysis. Six compounds were not taken in the model development.

The percentage inhibition for the compounds was observed to be zero at 3µg/ml. They are for compounds Cur_25, Cur_29, Cur_34, Cur_43, Cur_49, and Cur_56. The remaining compounds are fifty seven. This dataset was divided into, training set of forty six compounds (Cur_1, Cur_2, Cur_3, Cur_4, Cur_5, Cur_6, Cur_8, Cur_9, Cur_10, Cur_12, Cur_13, Cur_14, Cur_15, Cur_16, Cur_17, Cur_18, Cur_19, Cur_20, Cur_22, Cur_23, Cur_24, Cur_26, Cur_28, Cur_30, Cur_32, Cur_33, Cur_35, Cur_36, Cur_37, Cur_39, Cur_40, Cur_41, Cur_42, Cur_44, Cur_45, Cur_47, Cur_48, Cur_51, Cur_53, Cur_54, Cur_55, Cur_57, Cur_58, Cur_60, Cur_62, Cur_63) and the remaining eleven compounds as the test set (Cur_7, Cur_11, Cur_21, Cur_27, Cur_31, Cur_38, Cur_46, Cur_50, Cur_52, Cur_59 and Cur_61)

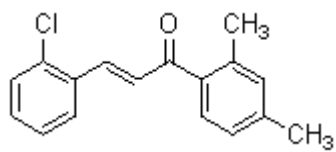
Table 7.1. Training set - forty six compounds

Cur_1 	Cur_2 	Cur_3 
Cur_4 	Cur_5 	Cur_6 

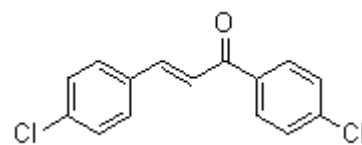
Cur_8

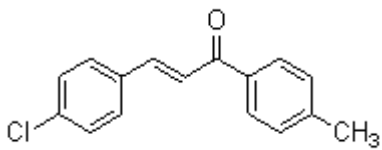
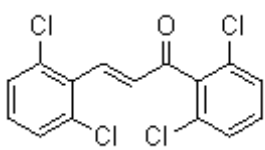
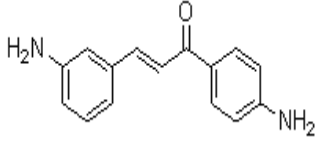
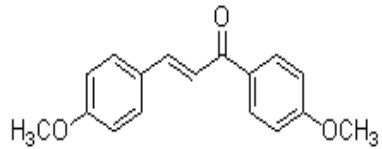
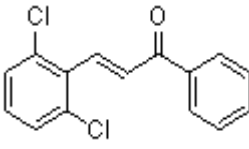
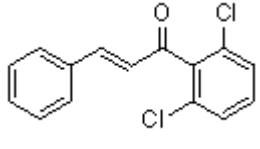
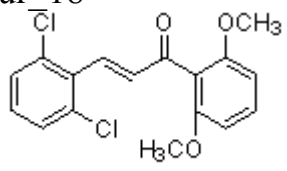
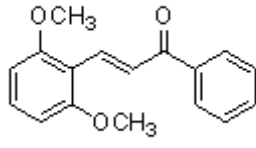
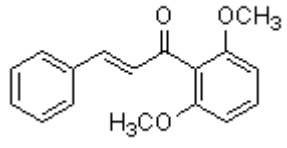
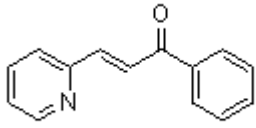
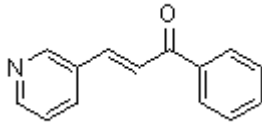
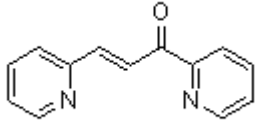
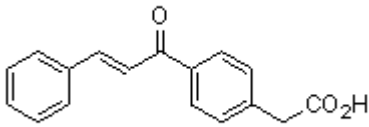
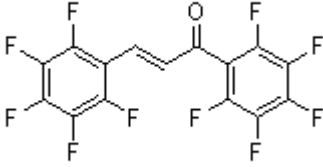
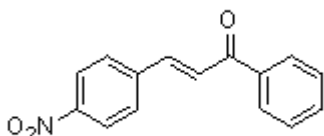
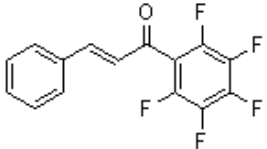
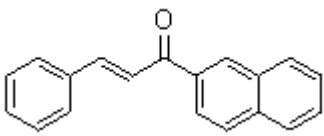
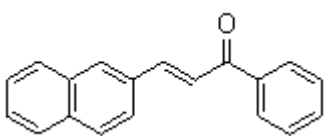


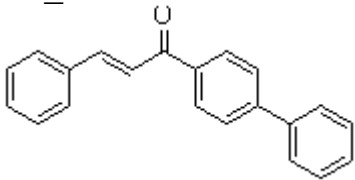
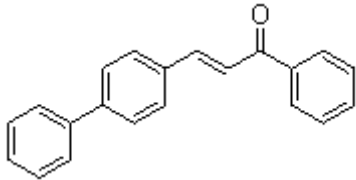
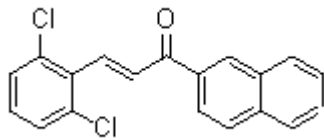
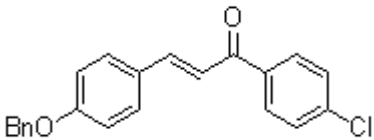
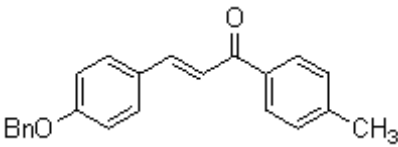
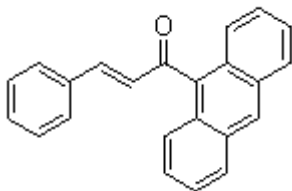
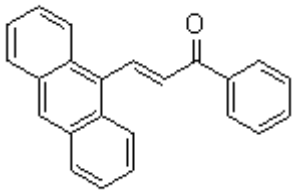
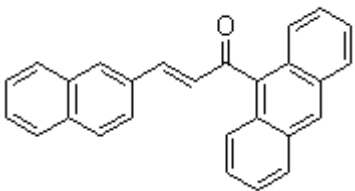
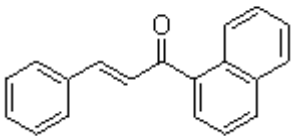
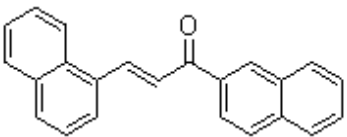
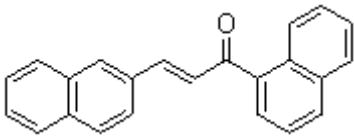
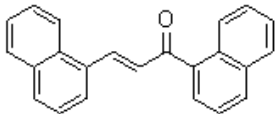
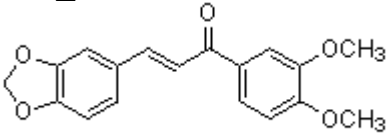
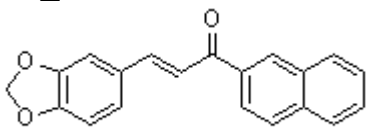
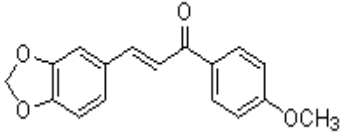
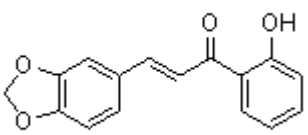
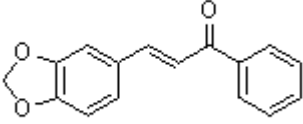
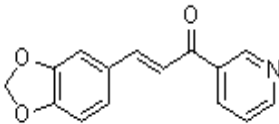
Cur_9



Cur_10



Cur_12 	Cur_13 	Cur_14 
Cur_15 	Cur_16 	Cur_17 
Cur_18 	Cur_19 	Cur_2 
Cur_22 	Cur_23 	Cur_24 
Cur_26 	Cur_27 	Cur_28 
Cur_30 	Cur_31 	Cur_32 

Cur_33 	Cur_35 	Cur_36 
Cur_37 	Cur_38 	Cur_39 
Cur_40 	Cur_41 	Cur_42 
Cur_44 	Cur_45 	Cur_46 
Cur_47 	Cur_48 	Cur_50 
Cur_51 	Cur_52 	Cur_53 

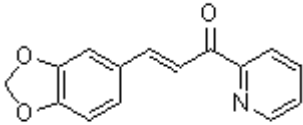
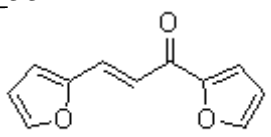
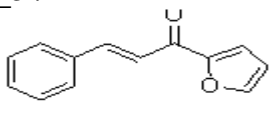
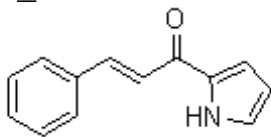
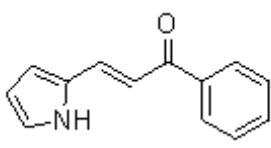
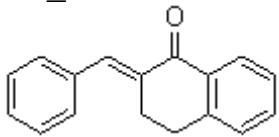
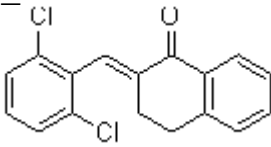
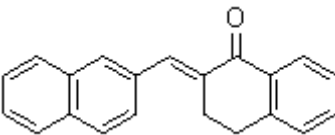
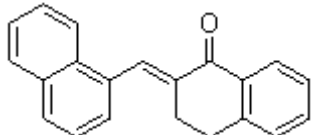
Cur_54	Cur_55	Cur_57
		
Cur_58	Cur_59	Cur_60
		
Cur_61	Cur_62	Cur_63
		

Table 7.2. Test set - Eleven compounds

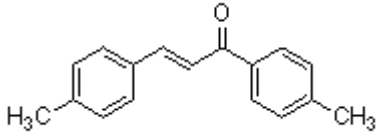
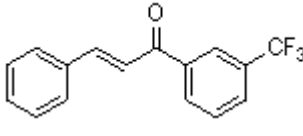
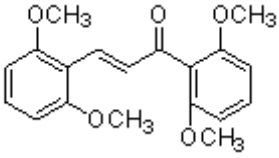
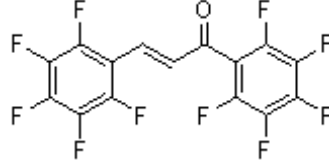
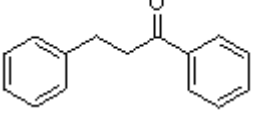
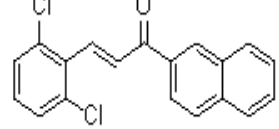
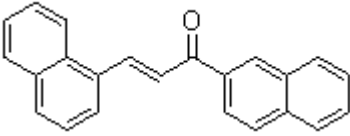
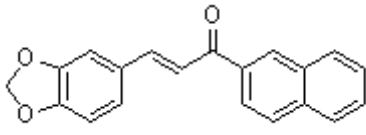
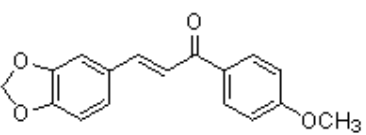
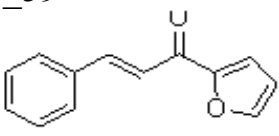
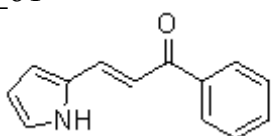
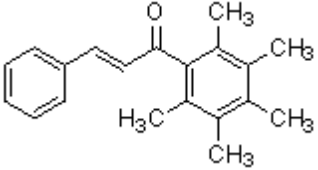
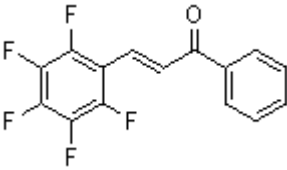
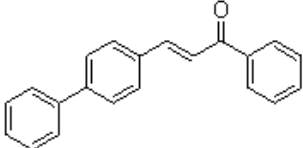
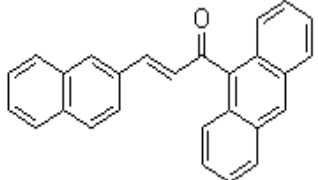
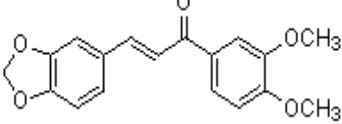
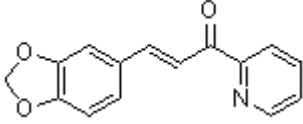
Cur_7 	Cur_11 	Cur_21 
Cur_27 	Cur_31 	Cur_38 
Cur_46 	Cur_50 	Cur_52 
Cur_59 	Cur_61 	

Table 7.3. Data set compounds not included in the QSAR study

Cur_25 	Cur_29 	Cur_34 
Cur_43 	Cur_49 	Cur_56 

Developing a CoMFA region file

Since the model yielded a low cross validated q^2 , rotation of the training set alignment was carried out in an effort to see if this might increase the predictive property of the model. The software creates the lattice and the lattice points based only on the initial alignment before the PLS run. This is termed as the CoMFA region file. The CoMFA region file has the dimensions in the x, y and z co-ordinates, the total number of lattice points, and the probe atom used to calculate the steric and electrostatic fields at the grid points. This CoMFA region file with lattice

points is only a few angstroms from the most distant atom from the center of the common alignment. As rotation of the alignment was carried out in the lattice, there is a possibility that the alignment shift beyond the PLS sampling lattice point, of the CoMFA region file. To avoid an error of not sampling fields, a customized lattice (new region file) was created which is large enough in all the three-dimensions. To create this CoMFA region file, a few molecules from the original database were rotated orthogonal (perpendicular) to the common alignment. These new molecules have new co-ordinates, and the molecules were renamed and were put in the database. The database has molecules with two different coordinates orthogonal to each other. A new molecular spread sheet was created with the above database. CoMFA columns were filled in the molecular spread sheet. The newly created CoMFA region file was saved. This saved CoMFA region file was used in the QSAR model after executing rotation and translation of the alignment of the training set.

Dimension of the original CoMFA region file, the region file is created automatically while creating the CoMFA columns in the molecular spread sheet

Points: 1320

Boxes: 1

Box 1	X	Y	Z
Lower Corner:	-11.974381	-14.146134	-9.157858
High Corner.:	9.540546	7.879578	8.835281
Step Size...:	2.000000	2.000000	2.000000
Number steps:	11	12	10
Probe Atom...:	C.3		
Charge.....:	1.000000		

Newly created CoMFA region file and its respective parameter

Points: 1755

Boxes: 1

Box 1	X	Y	Z
Lower Corner:	-16.250417	-12.516657	-8.289914
High Corner.:	13.068892	11.870170	8.756537
Step Size...:	2.000000	2.000000	2.000000
Number steps:	15	13	9
Probe Atom...:	C.3		
Charge.....:	1.000000		

In both the CoMFA region files, points indicate the number of lattice points. The lowest and higher corner indicate the total spacing in their respective dimensions (x, y and z coordinate axis) and the number of steps are the total steps in each co - ordinate. Probe atom C. 3 is the sp³

hybridized carbon used in calculating the fields at the grid points. Charge 1 indicates that the sp^3 hybridized carbon has a charge of +1

Table 7.4. PLS statistical results for CoMFA and CoMSIA 3D-QSAR model

PLS details	CoMFA	COMSIA
r^2	0.213	0.183
q^2	0.239	0.146
no validation r^2	0.965	0.946
S value	0.190	0.236
F value	127.682	81.055
Box Step size	2 Å	2 Å

Column filter of 2.0 kcal/mole was used during the PLS analysis in both the CoMFA and CoMSIA QSAR model development. Attenuation factor of 0.3 kcal/mole was used in the CoMSIA analysis.

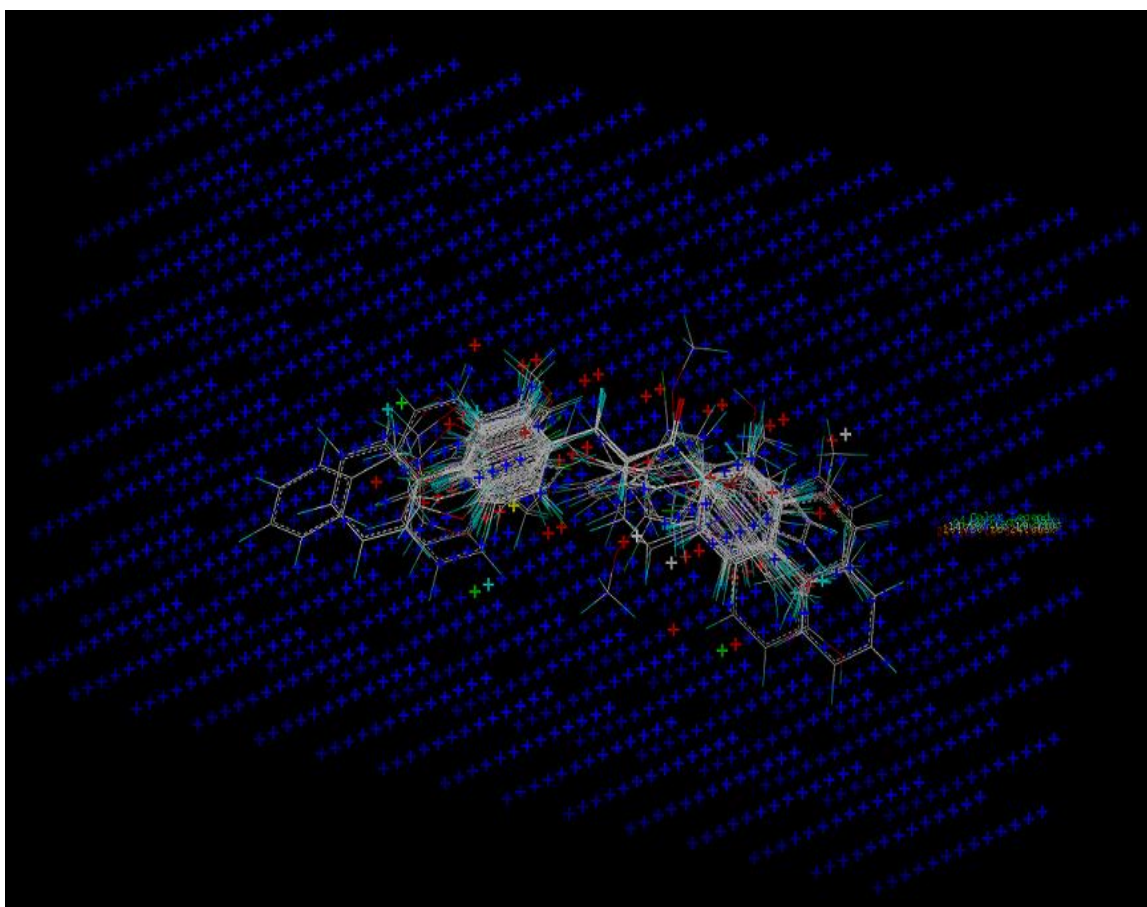


Fig 7.4 CoMFA region file with the loaded alignment, the blue dots indicate the lattice points of the region file.

Table 7.5. Results of the predictions of CoMFA and CoMSIA models

Compounds	Experimental	CoMFA prediction	CoMSIA Prediction	CoMFA Residuals	CoMSIA Residuals
Cur_20	3.68	3.66	3.41	0.02	0.27
Cur_1	4.53	4.09	4.15	0.44	0.38
Cur_10	2.46	2.98	2.81	-0.52	-0.35
Cur_11	4.47	3.45	4.61	1.02	-0.14
Cur_12	3.2	3.25	2.93	-0.05	0.27
Cur_13	4.32	4.28	4.10	0.04	0.22
Cur_14	3.64	3.75	3.64	-0.11	0.00
Cur_15	4.19	4.15	4.23	0.04	-0.04
Cur_16	4.11	4.13	4.36	-0.02	-0.25
Cur_17	3.93	3.99	3.92	-0.06	0.01
Cur_18	3.78	3.77	3.94	0.01	-0.16
Cur_19	4.03	4.03	4.00	0.00	0.03
Cur_2	3.69	3.27	3.42	0.42	0.27
Cur_21	3.90	3.34	3.62	0.55	0.28
Cur_22	4.49	4.61	4.58	-0.12	-0.09
Cur_23	3.46	3.52	3.45	-0.06	0.01
Cur_24	3.98	4.14	4.17	-0.16	-0.19
Cur_26	1.34	1.31	1.30	0.02	0.03
Cur_27	4.48	4.26	3.03	0.22	1.45
Cur_28	4.25	4.12	4.18	0.14	0.08
Cur_3	4.59	4.17	4.11	0.42	0.47
Cur_30	4.02	4.03	4.04	-0.01	-0.02
Cur_31	4.37	0.79	2.17	3.56	2.20
Cur_32	4.32	4.14	4.31	0.19	0.02

Cur_33	3.55	3.65	3.56	-0.10	-0.01
Cur_35	2.56	2.54	2.50	0.013	0.06
Cur_36	3.56	3.81	3.46	-0.25	0.11
Cur_37	1.48	1.35	1.38	0.13	0.11
Cur_38	1.63	1.56	1.50	0.07	0.13
Cur_39	3.38	3.87	3.92	-0.49	-0.53
Cur_4	2.34	2.51	2.55	-0.16	-0.22
Cur_40	3.98	3.94	4.21	0.04	-0.23
Cur_41	4.05	3.91	4.18	0.14	-0.13
Cur_42	3.50	3.61	3.400	-0.11	0.10
Cur_44	3.22	3.03	3.23	0.19	-0.02
Cur_45	3.11	3.15	2.90	-0.04	0.21
Cur_46	4.02	3.47	3.56	0.55	0.45
Cur_47	3.19	3.09	3.30	0.09	-0.11
Cur_48	0.37	0.37	0.41	-0.02	-0.04
Cur_5	2.65	2.86	3.16	-0.21	-0.51
Cur_50	2.67	1.11	1.92	1.56	0.75
Cur_51	2.54	2.52	2.62	0.02	-0.08
Cur_52	3.54	1.48	1.63	2.06	1.91
Cur_53	3.48	3.39	3.67	0.07	-0.18
Cur_54	3.79	3.97	3.91	-0.18	-0.12
Cur_55	2.09	2.07	1.90	0.02	0.19
Cur_57	4.05	4.11	4.18	-0.05	-0.12
Cur_58	2.43	2.40	2.44	0.03	-0.01
Cur_59	3.07	4.41	3.93	-1.34	-0.86
Cur_6	4.06	4.20	3.96	-0.14	0.10

Cur_69	2.43	2.26	2.50	0.18	-0.07
Cur_61	4.44	3.07	3.22	1.37	1.23
Cur_62	3.61	3.84	3.92	-0.23	-0.31
Cur_63	4.34	4.28	3.93	0.06	0.41
Cur_7	4.21	3.43	3.54	0.78	0.67
Cur_8	4.39	4.43	3.97	-0.04	0.42
Cur_9	3.23	3.21	3.59	0.02	-0.36

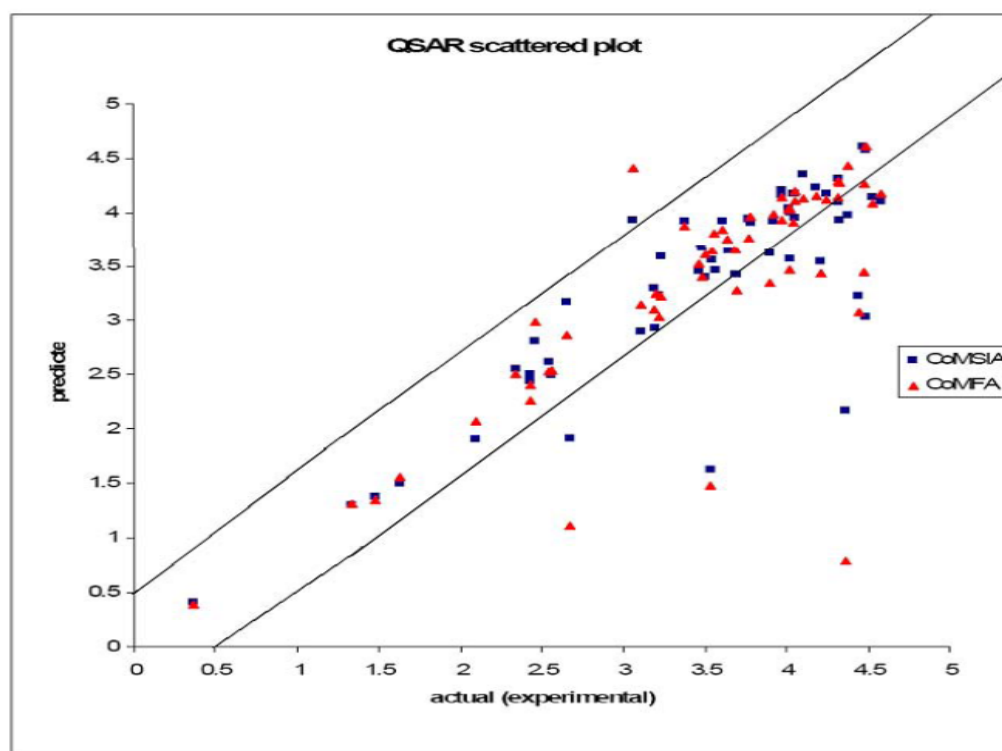


Fig 7.5 Scatter plot of CoMFA and CoMSIA analysis, predicted activity vs actual or experimental activity. The distribution of biological activity was observed to be in a range of five Log units. The points on the graph which are far away from the diagonal margin are considered to be outliers.

Affect of alignment on cross validated q^2 for the derived QSAR model - Rotation of alignment in the lattice

To observe how the cross validated q^2 is affected by the movement of the alignment of the training set compounds in the defined lattice; rotation of the alignment in the lattice was carried out. Rotation of the alignment with the x-axis as the center of rotation was performed. The STATIC ROTATE GLOBAL command was used via the Sybyl command line to execute this function in Sybyl. The rotation was carried out in increments of 20 degrees from 0 to 180 degrees in a clockwise direction. Since the lattice points of the CoMFA region file are orthogonal beyond 180 degrees, it is not necessary to carry additional rotation beyond 180 degrees because the PLS output will repeatedly be similar. The PLS results indicate there is a significant affect of the molecular alignment in the Cartesian grid on the cross validated q^2 . Such alignment affect is significantly low in CoMSIA as evident from the graph plotted, q^2 vs angle of rotation; see Table 7.6 and Figure 7.6.

Angle of Rotation	CoMFA q ²	CoMSIA q ²
0	0.289	0.14
20	0.197	0.13
40	0.169	0.16
60	0.128	0.101
80	0.224	0.103
100	0.212	0.115
120	0.174	0.134
140	0.234	0.094
160	0.239	0.062
180	0.106	0.148

Table 7.6. Cross validated q^2 observed in CoMFA and CoMSIA analysis with an incremental rotation in alignment.

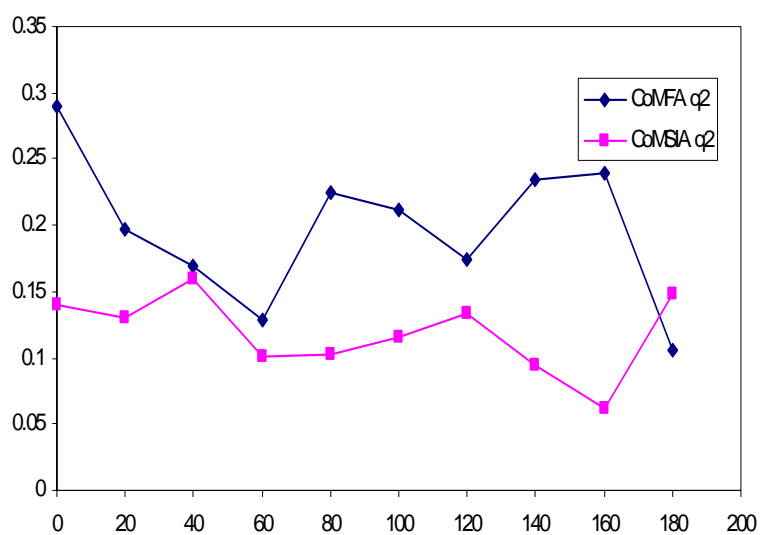


Fig 7.6 Cross validated q^2 vs angle of rotation, it is evident here q^2 value changes are minimum in CoMSIA analysis, where as in CoMFA analysis it is high.

CHAPTER VIII

DEPENDENT VARIABLE TRANSFORMATION

Most QSAR studies are carried out using a standard biological response (K_i , IC_{50} , LD_{50} , or ED_{50}). These biological responses are produced with a varied concentration of the compounds in their respective biological affinity assay (experimental activity). Here in the curcumin analogs dataset the case is different, a varied biological response (percentage inhibition) produces for a defined concentrations ($1\mu\text{g/ml}$, $3\mu\text{g/ml}$ and $6\mu\text{g/ml}$). In a effort to transform the dependent variable i.e percentage inhibition (ranging from 0 to 100 %) to inverse $\text{Log}_{10} ED_{50}$ ($\text{Log } 1/ED_{50}$), this function is also termed as pED_{50} . The below is the method used to determine the pED_{50} of the dataset. The derived three different pED_{50} calculated for the three different concentrations were used as dependent variables in the QSAR model with the dataset, even with their transformed variable. The predictions of the model were not acceptable; the observed crossvalidated q^2 observed was relatively low.

Essential of molecular pharmacology - The Saturation function²³

Determination of pED₅₀: Negative Log₁₀ ED₅₀

The determined pED₅₀ values were used as dependent variables in CoMFA and COMSIA studies for the curcumin analog dataset.

General form of the equation

$$\frac{y}{a} = \frac{x}{x+b} \quad 8.1$$

x - concentration of a drug, ligand or substrate.

y is a dependent variable, like amount of ligand bound, biological effect (IC₅₀, LD₅₀ or ED₅₀) or velocity in enzyme kinetics: Ki

a - maximum value that corresponds to y

When y is half maximal value then

$$\frac{y}{a} = 0.5 = \frac{x}{x+b}, \text{ so } x = b \quad 8.2$$

For Receptor – Ligand Binding

$$\frac{b}{B_{\max}} = \frac{[RD]}{[R]_T} = \frac{[D]}{[D] + K_D} \quad 8.3$$

K_D - dissociation constant; ED₅₀

[RD] - receptor drug complex.

$[R]_T$ - Total number of receptors.

When $K_D = [D]$

$$\frac{b}{B_{\max}} = 0.5 \quad 8.4$$

$$\text{Fractional biological response (effect)} = \frac{\% \text{ of maximal response}}{100} \quad 8.5$$

$$(\text{In our dataset, } f = \frac{\% \text{ Inhibition}}{100})$$

$$f = \frac{E}{E_{\max}}; E \text{ is percent inhibition and } E_{\max} \text{ is } 100 \%, \quad 8.6$$

$$\frac{E}{E_{\max}} \text{ is proportional to fraction of receptors occupied,}$$

$$\frac{b}{B_{\max}} = \frac{[RD]}{[R_T]} \quad 8.7$$

$$f = \frac{E}{E_{\max}} \text{ is proportional to } \alpha \frac{b}{B_{\max}} \quad (\alpha \text{ is a proportionality constant}).$$

If C = Drug conc.

$$\frac{b}{B_{\max}} = \frac{C}{C + K_D} \quad \text{In this study } C = \text{Molarity (M)} \quad 8.8$$

Molarity was calculated by converting the micrograms per milliliter of the drug solution to their respective Molar strength (Molarity - M)

$$f = \alpha \frac{C}{C + K_D} \quad 8.9$$

$$f K_D = \alpha C - f C \quad 8.10$$

$$f = C \left(\frac{\alpha - f}{K_D} \right)$$

$$K_D = C \frac{(\alpha - f)}{f}$$

Here α is a proportionality constant, if $\alpha = 1$

$$K_D = C \frac{(1 - f)}{f} \quad 8.11$$

K_D is the ED_{50}

Table 8.1. concentration C_1 , C_2 , and C_3 are the concentration expressed in

Molar strength: Molarity calculated for 1 $\mu\text{g/ml}$, 3 $\mu\text{g/ml}$ and 6 $\mu\text{g/ml}$.

percentage inhibitions are indicated as P_1 , P_2 and P_3 respectively.

Molecule name	Molecular weight	C_2	C_1	C_3	P_1	P_2	P_3
Cur_7	236.31	12.70	4.24	25.39	36.3	67.3	89.5
Cur_61	284.36	10.55	3.52	21.10	84.7	84.9	78.8
Cur_3	291.18	10.30	3.43	20.61	73.7	98.2	98.1
Cur_1	208.26	14.40	4.80	28.81	71.6	92.8	94.4
Cur_40	308.38	9.73	3.24	19.46	54.6	53.5	71.3
Cur_41	358.44	8.37	2.79	16.74	52.5	57.3	48.1
Cur_13	346.04	8.67	2.89	17.34	48.3	75.3	93.7
Cur_6	236.31	12.69	4.23	25.39	47.7	57.9	89.6
Cur_28	253.26	11.85	3.95	23.69	47.6	70.3	84.5
Cur_14	238.29	12.59	4.20	25.18	43.5	38.1	69.1
Cur_11	276.26	10.86	3.62	21.72	42.3	87.4	96.9
Cur_8	236.31	12.69	4.23	25.39	41.7	80.4	87.3
Cur_22	209.25	14.34	4.78	28.67	40.7	89.1	96.9
Cur_51	268.27	11.18	3.73	22.37	36.5	12.7	24.5
Cur_21	328.36	9.14	3.04	18.27	36.2	49.2	39.2
Cur_53	253.26	11.85	3.95	23.69	33	32.5	68.9
Cur_62	303.19	9.89	3.30	19.79	32.8	37	53.7
Cur_19	268.31	11.18	3.73	22.36	31.8	56.4	60.8
Cur_27	388.16	7.73	2.58	15.46	31.1	88.6	88.6
Cur_52	252.27	11.89	3.96	23.78	31	34.3	80.8

Cur_23	209.25	14.34	4.78	28.67	30.3	31.7	83.2
Cur_2	256.73	11.68	3.89	23.37	29.9	40.2	58.5
Cur_9	270.76	11.08	3.69	22.16	29.6	25.3	73.4
Cur_15	268.31	11.18	3.73	22.36	29.1	63.4	85.2
Cur_54	253.26	11.85	3.95	23.69	28.3	44.1	78.4
Cur_57	198.22	15.13	5.04	30.27	28.2	57.7	90.2
Cur_20	268.31	11.18	3.73	22.36	25.8	39.8	63.5
Cur_37	348.83	8.60	2.87	17.20	25.6	4.4	62.9
Cur_39	308.38	9.73	3.24	19.46	25.5	46.4	69
Cur_35	284.36	10.55	3.52	21.10	24.9	12.9	41.4
Cur_18	337.20	8.90	2.97	17.79	23.2	43.7	52.3
Cur_12	256.73	11.68	3.89	23.37	22.3	24.5	46.5
Cur_49	266.30	11.27	3.75	22.53	21.3	0.1	0.1
Cur_4	291.18	10.30	3.43	20.61	19.9	10.4	84.7
Cur_5	264.37	11.35	3.78	22.70	19.5	14.2	59.2
Cur_34	360.46	8.32	2.77	16.65	19.2	0.1	0.1
Cur_59	197.24	15.21	5.07	30.42	19	21.5	0.4
Cur_38	328.41	9.14	3.04	18.27	18.4	5.1	17
Cur_10	277.15	10.82	3.61	21.65	13.8	11.7	31.1
Cur_26	266.30	11.27	3.75	22.53	13.7	3.8	29.7
Cur_25	278.39	10.78	3.59	21.55	12.6	0.1	5.3
Cur_36	327.21	9.17	3.06	18.34	12.2	35.3	96.8
Cur_46	308.38	9.73	3.24	19.46	12	55.5	60.3
Cur_42	258.32	11.61	3.87	23.23	10.9	33.2	68.2
Cur_33	284.36	10.55	3.52	21.10	10.4	34.7	0.1
Cur_56	198.22	15.13	5.04	30.27	9.8	1.0	0.1

Cur_58	197.24	15.21	5.07	30.42	8.3	11.4	19.4
Cur_55	188.18	15.94	5.314	31.88	7.7	8.1	12
Cur_30	298.21	10.060	3.35	20.12	6.6	55.6	88.7
Cur_48	302.33	9.92	3.31	19.85	5.7	1.4	11.9
Cur_60	234.30	12.80	4.27	25.61	5.7	11.4	40.8
Cur_16	277.15	10.82	3.61	21.65	4.6	61	94
Cur_17	277.15	10.82	3.61	21.65	2.2	51.1	88.7
Cur_47	312.32	9.61	3.202	19.211	1.9	24.2	48.1
Cur_24	210.24	14.27	4.76	28.54	0.1	53.4	85.2
Cur_29	298.21	10.06	3.35	20.12	0.1	0.1	0.1
Cur_31	258.32	11.61	3.87	23.23	0.1	78.7	88.6
Cur_32	258.32	11.613	3.87	23.23	0.1	75.4	66.2
Cur_43	408.5	7.344	2.45	14.69	0.1	0.1	4.7
Cur_44	308.38	9.73	3.24	19.46	0.1	24.9	50.3
Cur_45	308.38	9.73	3.24	19.46	0.1	22.4	52
Cur_50	282.29	10.63	3.54	21.25	1	14.4	45.5
Cur_63	284.36	10.55	3.52	21.10	1	76.4	52.1

Table 8.1. f_1 , f_2 and f_3 and the derived pED_{50} for the three different concentrations.

$pED_{50}-1$: pED_{50} determined for $1\mu\text{g/ml}$

$pED_{50}-2$: pED_{50} determined for $3\mu\text{g/ml}$

$pED_{50}-3$: pED_{50} determined for $6\mu\text{g/ml}$

STD: standard deviation

Average pED_{50} : Mean of $pED_{50}-1$, $pED_{50}-2$ and $pED_{50}-3$

f_1	f_2	f_3	$pED_{50}-1$	$pED_{50}-2$	$pED_{50}-3$	STD	Average pED_{50}
0.36	0.67	0.89	5.13	5.21	5.530	0.21	5.29
0.85	0.85	0.79	6.20	5.73	5.25	0.48	5.72
0.74	0.98	0.98	5.91	6.72	6.40	0.42	6.34
0.72	0.93	0.94	5.72	5.95	5.77	0.12	5.81
0.56	0.53	0.71	5.570	5.07	5.11	0.28	5.25
0.52	0.57	0.48	5.60	5.20	4.74	0.43	5.18
0.48	0.75	0.94	5.51	5.55	5.93	0.23	5.66
0.48	0.58	0.90	5.33	5.03	5.53	0.25	5.30
0.48	0.70	0.84	5.36	5.30	5.36	0.03	5.34
0.43	0.38	0.69	5.26	4.69	4.95	0.29	4.97
0.42	0.87	0.97	5.31	5.80	6.19	0.43	5.76
0.42	0.80	0.87	5.23	5.51	5.43	0.15	5.39
0.41	0.89	0.97	5.16	5.76	6.04	0.45	5.65
0.36	0.13	0.24	5.19	4.11	4.16	0.61	4.49
0.36	0.49	0.39	5.27	5.025	4.55	0.37	4.95
0.33	0.32	0.69	5.10	4.61	4.97	0.25	4.89
0.328	0.37	0.54	5.17	4.77	4.77	0.23	4.90

0.32	0.56	0.61	5.10	5.06	4.84	0.14	5.00
0.31	0.89	0.89	5.24	6.00	5.70	0.38	5.65
0.31	0.34	0.81	5.05	4.64	5.25	0.31	4.98
0.30	0.32	0.83	4.96	4.51	5.24	0.37	4.90
0.299	0.40	0.58	5.04	4.76	4.78	0.16	4.86
0.30	0.25	0.73	5.06	4.48	5.09	0.34	4.88
0.29	0.63	0.85	5.04	5.19	5.41	0.19	5.21
0.28	0.44	0.78	4.00	4.82	5.18	0.18	5.00
0.28	0.58	0.90	4.89	4.95	5.48	0.32	5.11
0.26	0.40	0.63	4.97	4.77	4.89	0.10	4.88
0.26	0.04	0.63	5.08	3.73	4.99	0.76	4.60
0.25	0.46	0.70	5.02	4.95	5.06	0.06	5.01
0.25	0.13	0.41	4.97	4.15	4.52	0.41	4.55
0.23	0.44	0.52	5.01	4.94	4.79	0.11	4.91
0.22	0.24	0.46	4.87	4.44	4.57	0.22	4.63
0.21	0.00	0.00	4.86	1.95	1.65	1.77	2.82
0.20	0.10	0.85	4.86	4.05	5.43	0.69	4.78
0.19	0.14	0.59	4.81	4.16	4.82	0.37	4.59
0.19	0.00	0.00	4.93	2.08	1.78	1.74	2.93
0.19	0.21	0.00	4.66	4.25	2.12	1.37	3.68
0.18	0.05	0.17	4.87	3.77	4.05	0.57	4.23
0.14	0.12	0.31	4.65	4.09	4.32	0.28	4.35
0.14	0.04	0.30	4.63	3.54	4.27	0.55	4.15
0.13	0.00	0.05	4.60	1.97	3.41	1.32	3.33
0.12	0.35	0.97	4.66	4.77	6.22	0.87	5.22
0.12	0.55	0.60	4.62	5.11	4.89	0.24	4.87

0.11	0.33	0.68	4.50	4.63	4.96	0.24	4.70
0.10	0.35	0.00	4.52	4.70	1.68	1.70	3.63
0.10	0.0	0.00	4.33	2.82	1.52	1.41	2.89
0.08	0.11	0.19	4.25	3.93	3.90	0.20	4.03
0.08	0.08	0.12	4.20	3.74	3.63	0.30	3.86
0.07	0.57	0.89	4.32	5.09	5.59	0.64	5.00
0.06	0.014	0.12	4.26	3.16	3.83	0.56	3.75
0.06	0.11	0.41	4.15	4.00	4.43	0.22	4.19
0.05	0.61	0.94	4.13	5.16	5.86	0.87	5.05
0.02	0.51	0.89	3.79	4.98	5.56	0.90	4.78
0.02	0.24	0.48	3.78	4.52	4.68	0.48	4.33
0.00	0.53	0.85	2.32	4.90	5.30	1.62	4.19
0.00	0.00	0.00	2.47	2.00	1.70	0.39	2.06
0.00	0.79	0.89	2.41	5.50	5.52	1.79	4.48
0.00	0.75	0.66	2.41	5.42	4.93	1.61	4.25
0.00	0.00	0.05	2.61	2.13	3.53	0.71	2.76
0.00	0.25	0.50	2.49	4.53	4.72	1.24	3.91
0.00	0.22	0.52	2.49	4.47	4.75	1.23	3.90
0.01	0.14	0.45	3.45	4.20	4.59	0.58	4.08
0.01	0.76	0.52	3.46	5.49	4.71	1.02	4.55

CHAPTER IX

RESULTS AND DISCUSSIONS

Curcumin and several of its structural analogs have a diverse range of activity towards various mediators of cancer progression. Selectivity is an important factor. Initial pharmacophore perception in MOE (Molecular Operating Environment) results in the identification of the interaction pharmacophore elements: IPEs. There are two aromatic systems: F₃-Ar₁ and F₄-Ar₂ as well as a hydrogen bond donor - F₁ and a hydrogen bond acceptor - F₂, see figure 9.1. The pharmacophore model developed is a four point pharmacophore model. These interacting pharmacophore elements (F₁, F₂, F₃ and F₄) are imperative for biological activity. The spacing of the pharmacophore elements was indicated in the Figure 9.1. This pharmacophore was used as a search query in the screening of the National Cancer Institute (NCI) database of 21607 compounds. After successfully screening the NCI database, 1471 compounds were identified as hits. These hits can be further filtered with Lipinski's Rule of Five.²⁴ The rule determines the essential features for a drug-like molecule, even though a molecule might be highly potent if it violates the Lipinski's Rule of Five, it is unlikely, but not absolutely

guaranteed, that the molecule cannot be developed to a future clinical candidate.

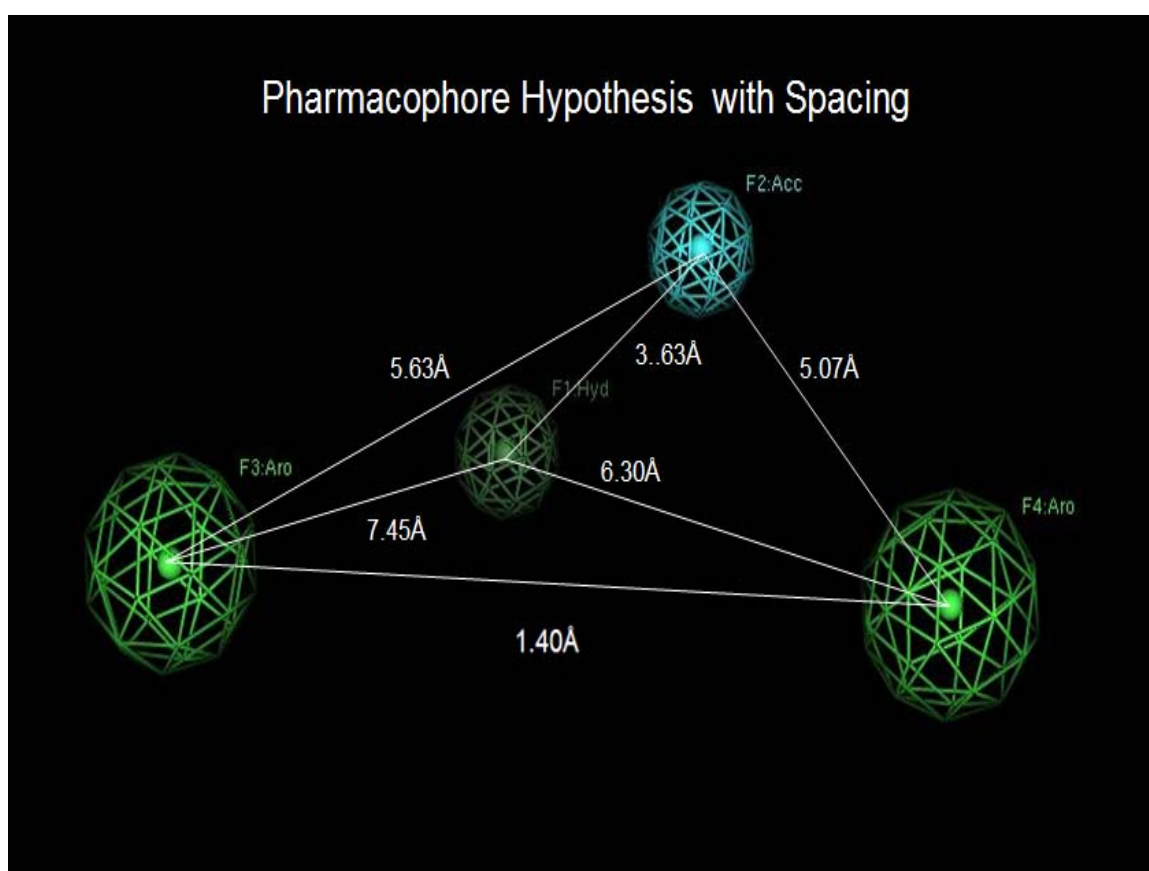


Fig 9.1. Four point pharmacophore hypothesis of aromatic enones with spacing between the pharmacophore elements. F₃ and F₄ are the aromatic systems, F₂ is the Hydrogen bond donor and F₁ is the hydrogen bond acceptor.

Information related to biological target specific binding of curcumin analogs is lacking. Thus, ligand based modeling is the best available option. Three-dimensional QSAR analysis is one of the most reliable methods in ligand based computational model development.

Initially when all the sixty-three structures were taken collectively in model building, the statistical results of the PLS runs were not good. The convergent r^2 was approximately in a range of 0.08 magnitude of the fifty seven compound dataset. When grouping the dataset into the training and test sets, no correlation was identified in the training set between the molecular field descriptors and the dependent variable (natural log function of percentage inhibition at 3 μ g/ml concentration).

Usually three-dimensional QSAR studies are carried out with standard biological response (IC₅₀, LD₅₀, ED₅₀, Ki - which is converted to an inverse Log function) observed at a defined concentration of each compound in the cell based assay study. In the present biological assay it is different, a varied biological response (percentage growth inhibition) at a defined concentration (three different concentrations). So a pharmacodynamic conversion was derived with the available two variables (percentage inhibition and concentration of the drug solution)

into ED₅₀ and the negative log function (pED₅₀) is taken as the dependent variable in the QSAR model development. Even with effort an acceptable correlation was achieved with the model. When six compounds Cur_25, Cur_29, Cur_34, Cur_43, Cur_49, and Cur_56 were eliminated from the training set, there is a correlation observed in the model. The crossvalidated q^2 for CoMFA was 0.289 and 0.146 in CoMSIA analysis. The results and the outcomes of this final QSAR model are acceptable, which is evident in the final validation of the model with the prediction set compounds and the contour maps. The feature present in the molecule complements the QSAR contour maps of the model. Standard deviation method is used in deriving the contour maps.

A few compounds with good prediction in the training set and the test were placed in the contour maps. The predictions of the derived three- dimensional QSAR model were acceptable. When few compounds of the training set and the test set were chosen to validate the model, the molecular features were observed. By visualizing them in the QSAR contour maps, there is a compatibility observed in the predictions. Therefore, with structural modification with different functional groups and their respective orientation in three-dimensional space might result in increased activity of the molecules as evident with the contour maps.

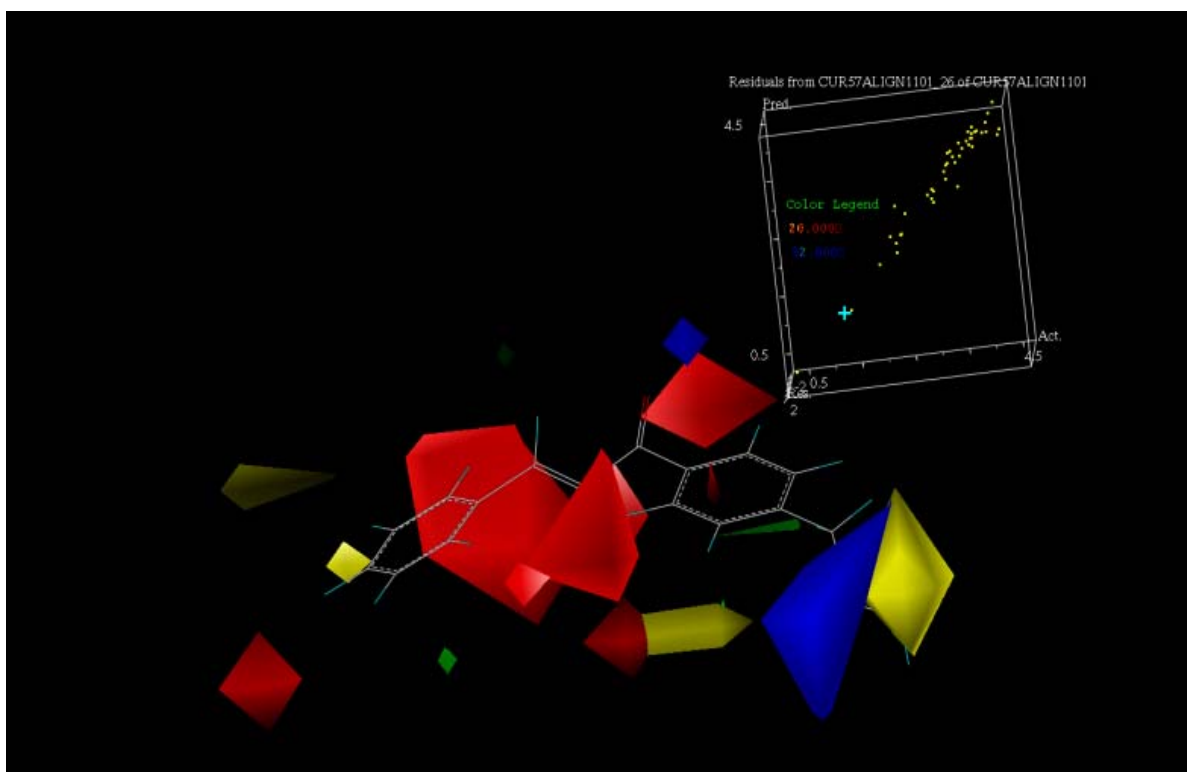


Fig 9.2. CoMFA stereo view contour coefficient maps, Cur_26, is placed in the three dimension contour map.

CoMFA contour maps color codes

Interaction fields	Green	Yellow
Steric	Favors interaction	Disfavors interaction
Electrostatic	Blue favors positively charged group interaction	Red favors negatively charged group interaction

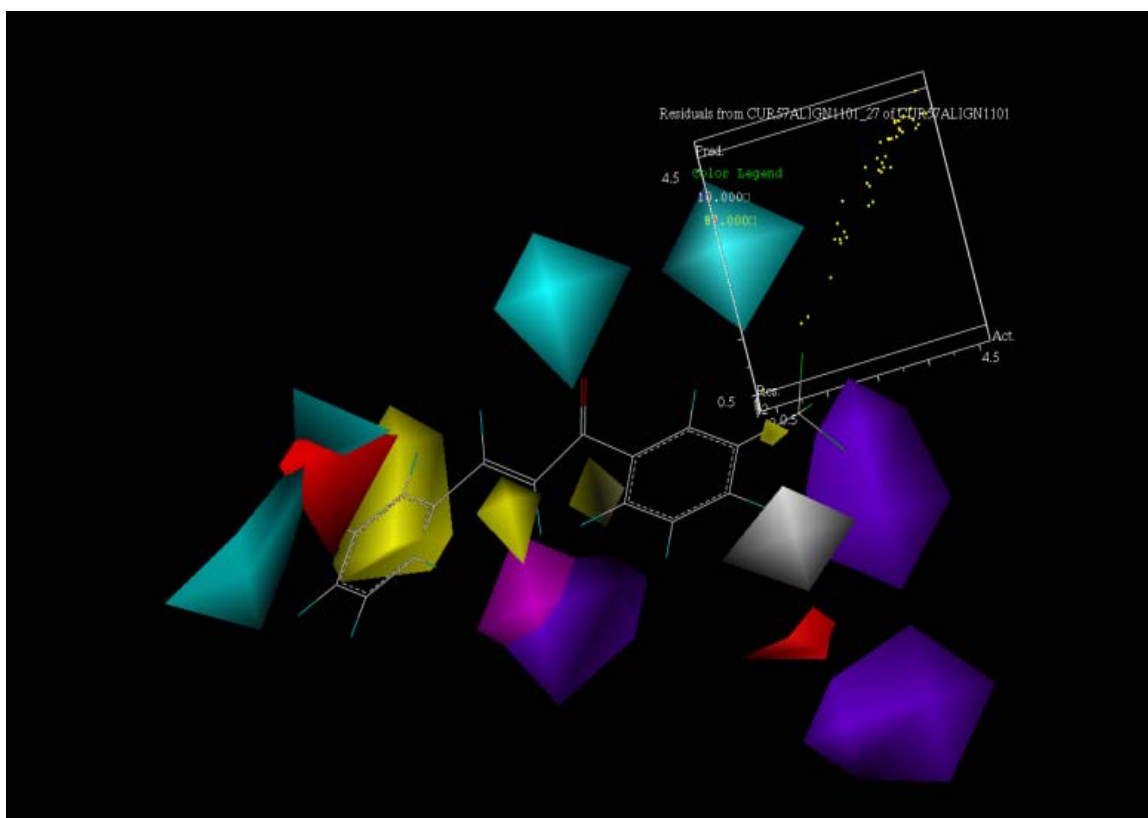


Figure 9.3. CoMSIA stereo view, the upper right corner has the scatter plot of the training set compounds Cur_11, a test set compound placed in the CoMSIA contour map.

CoMSIA contour coefficient map color codes

Interaction Fields	Favor interaction	Disfavor interaction
Steric	Green	Yellow
Electrostatic	Blue	Red
Hydrophobicity	Yellow	White
Hydrogen bond acceptor	Magenta	Red
Hydrogen bond donor	Cyan	Purple

The contour maps shown below indicate the individual CoMSIA field contributions and their corresponding stereo view. Some of the training set and test set compounds were placed in the contour coefficient maps for qualitative inspection for identification of various groups that favor and disfavor interaction and their contribution towards activity.

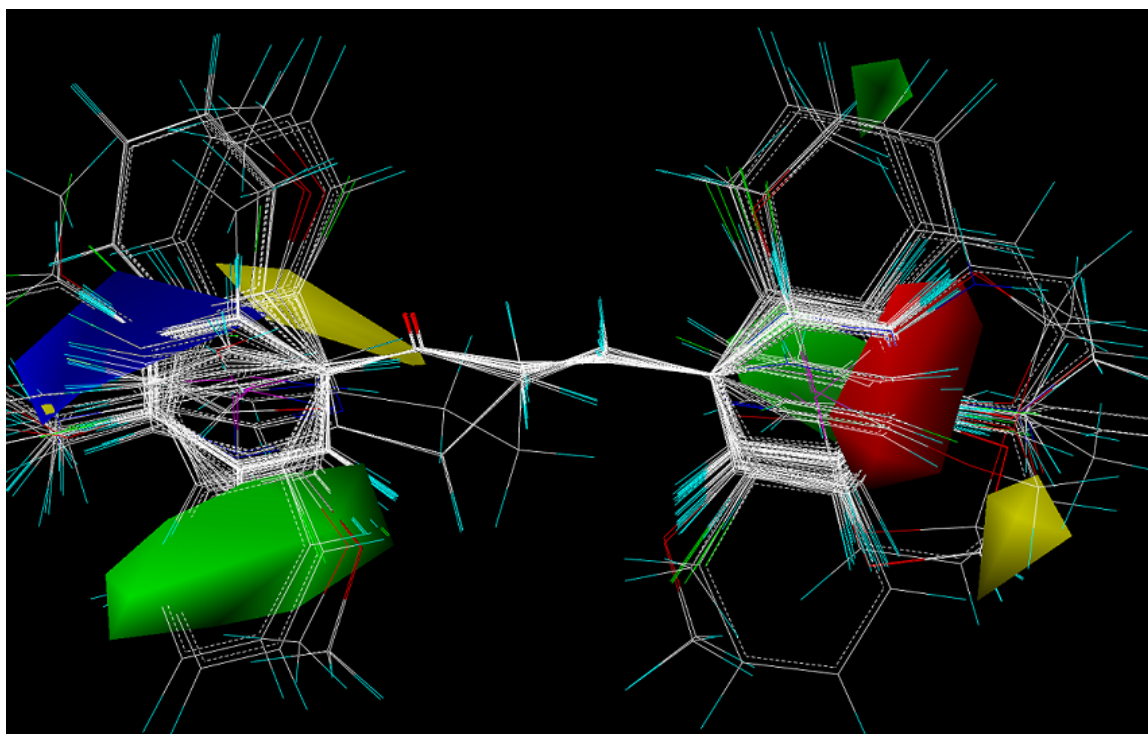
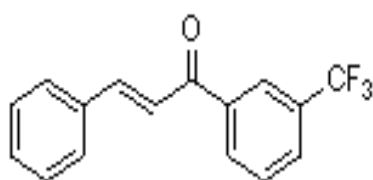
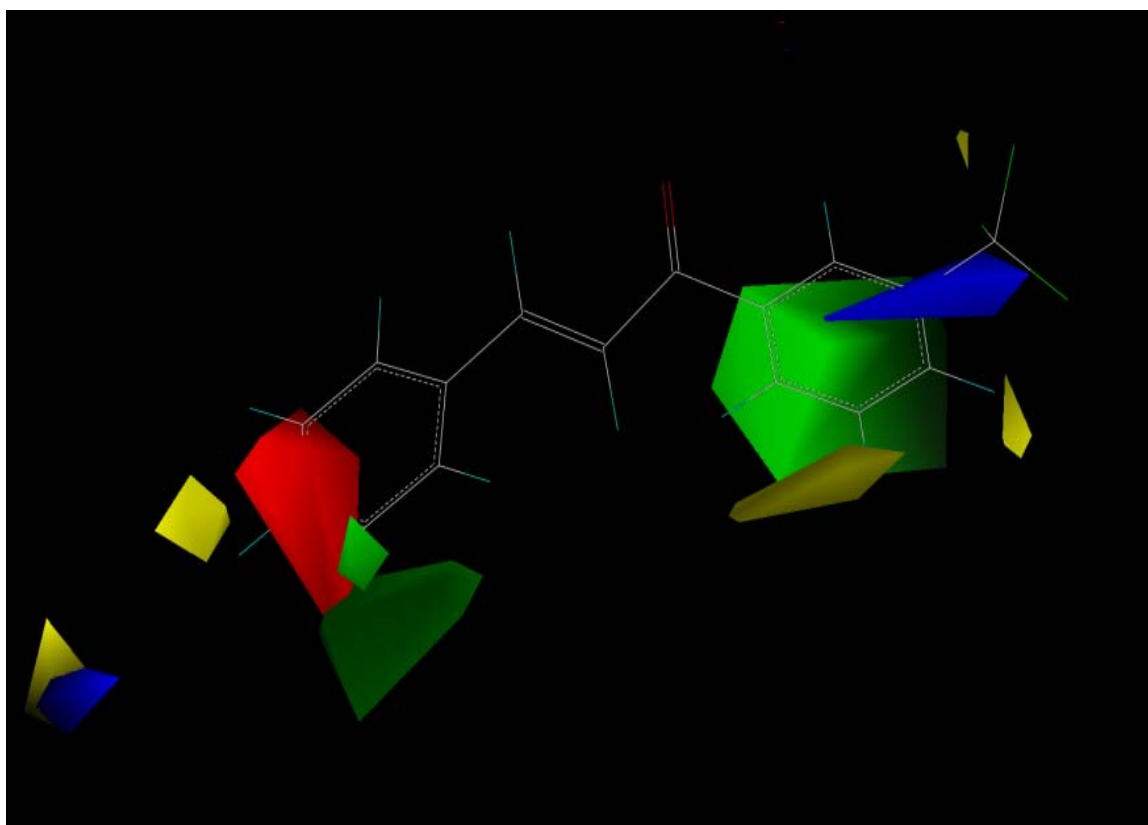


Fig 9.4. CoMSIA steric and electrostatic contour map with the total dataset alignment.

Green indicates steric interactions are favorable in the regions and yellow indicates it is disfavored. For electrostatic interactions, blue indicates positively charged groups are favored in the region and red indicates negatively charged groups are favored.

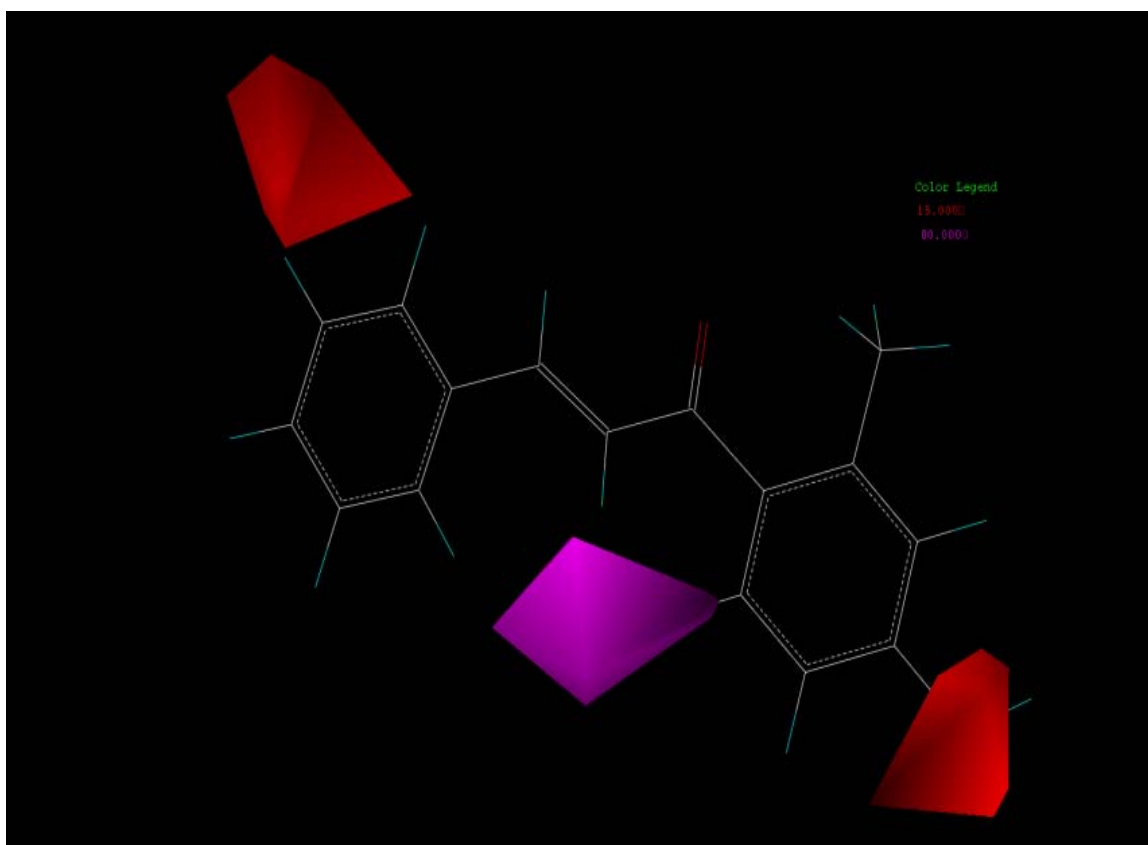


Contour details
Steric green – 92.00 and yellow – 20.00
Electrostatic blue – 91.00 and red – 18.00

Fig 9.5. Steric and Electrostatic contour maps. Training set compound cur_11

Steric: Green indicates steric interactions are favored in the region and yellow indicates it is disfavored.

Electrostatic: Blue indicates positively charged electrostatic interactions are favored in that region and red indicates it is disfavored. In the map there is a trifluoro group present on R₂ aromatic system, which is a strong prediction of the model.



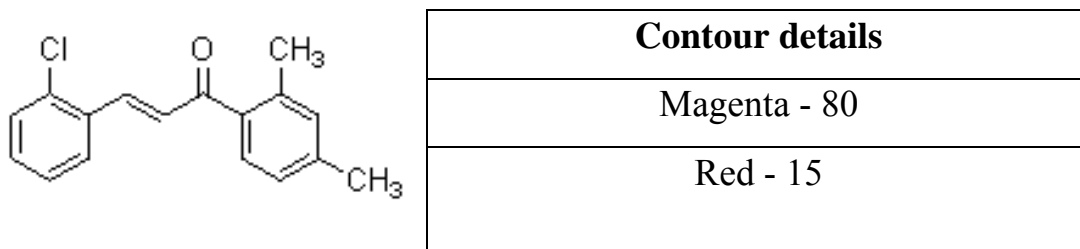
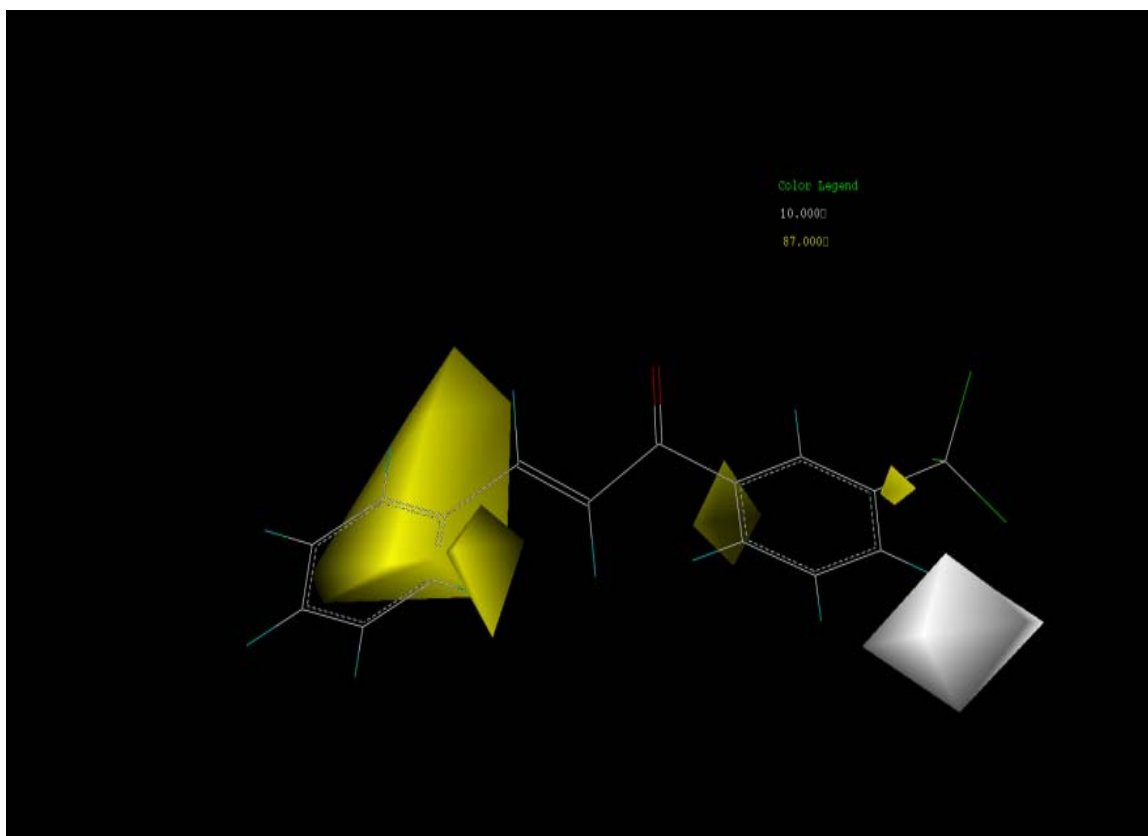


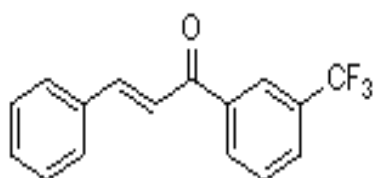
Fig 9.6. Training set compound cur_9

HYDROGEN BOND ACCEPTOR SYSTEM

Magenta indicates H-bond acceptor groups are favored in the region of the molecule. Magenta - 80

Red indicates H-bond acceptor groups are disfavored in that region.

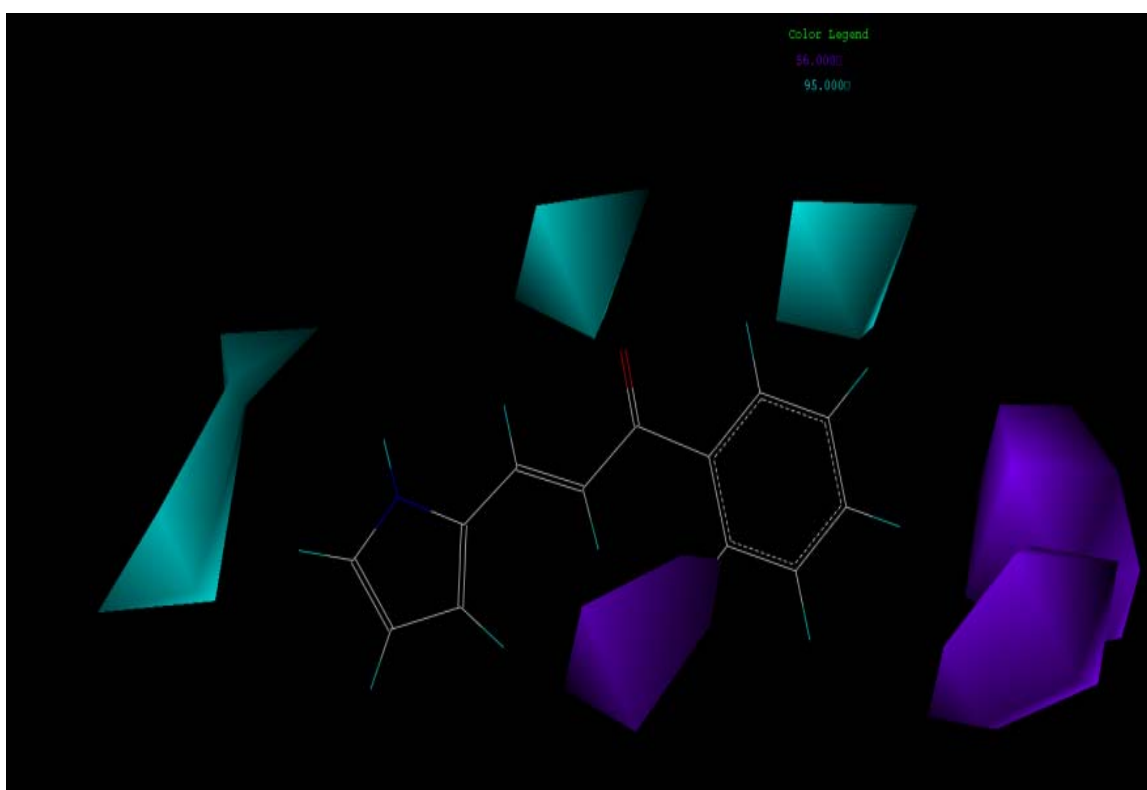


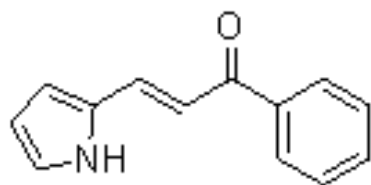


Contour details
Yellow – 87.00
White – 10.00

Figure 9.7. Test set compound Cur_11

Hydrophobicity: yellow indicates hydrophobic group are favored in the region and white indicates it is disfavored.





Contour details
Cyan – 95.00
Purple – 56.00

Fig 9.8. Training set compound cur_59

Hydrogen bond donor system

Cyan indicates hydrogen bond donor groups are favored in that region and purple indicates hydrogen bond donor groups are disfavored in the region.

CHAPTER X

FUTURE RESEARCH

Curcumin is considered as an important lead candidate in biochemical research as evident from various sources. As its activity profile is not unique to only a specific target or to a class of closely related biological targets, future research should be focused to identify all the targets to which curcumin exhibits its affinity/inhibition potential. *In silico* and computational research methods can be feasible if the structural data of the targets are reported. These are usually reported as a co-crystal structure of the biological target with the ligand bound to the active domain or a NMR data of the putative target. The ligand can be curcumin or compounds that have some degree of structural resemblance with curcumin. Solubility is a major hurdle for the development of curcumin or its structural analogs. These factors need to be taken into account in future research that focus on the design and synthesis of compounds that have greater water solubility.

BIBLIOGRAPHY

- 1) Furness, M. S.; Robinson, T. P.; Ehlers, T.; Hubbard, R. B., IV; Arbiser, J. L.; Goldsmith, D. J.; Bowen, J. P. *Current Pharmaceutical Design* **2005**, 11(3), 357.
- 2) Pharmacological basis of therapeutics Goodman Gilman.
- 3) Text book of pharmaceutical and medicinal chemistry., Wilson and Gisvold.
- 4) Carmeliet, P.; Rakesh, K.J. Angiogenesis in cancer and other diseases. *Nature*, **2000**, 407, 249-257.
- 5) Folkman, J. Tumor angiogenesis-therapeutic implications. *New England Journal of Medicine*, **1971**, 285, 1182-1186.
- 6) Arbiser, J. L.; Klauber, N.; Rohan, R.; van Leeuwen, R.; Huang, M. T.; Fisher, C.; Flynn, E.; Byers, H. R. *Molecular Medicine* **1998**, 4, 376.
- 7) Salimath, B. P. Molecular mechanisms of anti-angiogenic effects of curcumin. *Biochemical and Biophysical Research communications* **2002**, 297, 934-942.
- 8) Leu T.H.; Maa M. C.; The molecular mechanisms for the antitumorigenic effects of curcumin. *Current medicinal chemistry anticancer agents* **2002**, 2(3):357-70.
- 9) Arbiser, J. L.; Moses, M. A.; Fernandez, C. A.; Ghiso, N.; Cao, Y. H.; Klauber, N.; Frank, D.; Brownlee, M.; Flynn, E.; Parangi, S.; Byers, H. R.; Folkman, J. *Proc.Natl. Acad. Sci.* **1997**, 94, 861-866.

- 10) Robinson, T. P.; Ehlers, T.; Hubbard, R. B.; Bai, X. H.; Arbiser, J. L.; Goldsmith, D. J.; Bowen, J. P. *Bioorg. Med. Chem. Lett.* **2003**, 13(1): 115.
- 11) Fujita, T.; Iwasa, J.; Hansch, C. J. *Am. Chem. Soc.* **1964**, 86 5175.
- 12) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, 110, 5959-596.
- 13) G. Klebe, U. Abraham, and T. Mietzner, *J. Med. Chem.*, **1994**, 37, 4130-4146.
- 14) M. E. Wolff, (ed.), 'Burger's Medicinal Chemistry', **Vol. I: Principles and Practice**, Wiley, New York, 5th edn., 1995.
- 15) Paul Geladi., Bruce R Kowalski., *Analytica Chimica Acta.* **1986**, 186, 1-1
- 16) Arbiser, J. L.; Moses, M. A.; Fernandez, C. A.; Ghiso, N.; Cao, Y.; Klauber, N.; Frank, D.; Brownlee, M.; Flynn, E.; Parangi, S.; Byers, H. R.; Folkman, J. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, 94(3), 861.
- 17) H. Kubinyi, (ed.), '3D QSAR in Drug Design. Theory, Methods and Applications', ESCOM, Leiden, 1993.
- 18) Dissertation.; Tedman J. Ehlers Development of antiangiogenic compounds: computational studies of curcumin based drugs and methaminopeptidase type II inhibitors.
- 19) Thomas Philip Robinson, Richard B. Hubbard, IV, Tedman J. Ehlers, Jack L. Arbiser, David J. Goldsmith and J. Phillip Bowen, *Bioorganic & Medicinal Chemistry*, **2005**, 13 4007–4013.

- 20) Tripos Inc. SYBYL molecular modeling software; Tripos Inc., 1699 South Hanley Rd, Suite 303, St. Louis, MO 63144.
- 21) Tripos force Fields., vinter J. G.; Davis, A.; Saunder, M. R. *J. Comp-Aided Mol. Design*, **1987**, *1*, 31.
- 22) Gyanendra Pandey, Anil K. Saxena *J. Chem. Inf. Model.* **2006**, *46*, 2579-2590.
- 23) Receptors in a Quantitative approach, Alexander levitziski.
- 24) Christopher A Lipinski, Franco Lombado, Beryl W. Dominy, Paul J. Feeney, *Advanced Drug Delivery Reviews*, **2001**, *46*, 3 - 26.